SIGNATRON®

AFCRL-69-0256

AD 690179

INVESTIGATION OF FACTORS AFFECTING THE QUALITY OF VOCODER SPEECH

by

Thomas H. Crystal
SIGNATRON, Inc., 594 Marrett Road, Lexington, Massachusetts 02173

Contract No. F19628-67-C-0292
Project No. 4610
Task No. 461002
Unit No. 46100201

FINAL REPORT

Period Covered: April 15, 1967 through May 17, 1969

May 17, 1969

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Prepared
for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

82

SIGNATRON®

AFCRL-69-0256

INVESTIGATION OF FACTORS AFFECTING THE QUALITY OF VOCODER SPEECH

by

Thomas H. Crystal

SIGNATRON, Inc., 594 Marrett Road, Lexington, Massachusetts 02173

Contract No. F19628-67-C-0292

Project No. 4610
Task No. 461002
Unit No. 46100201

FINAL REPORT

Period Covered: April 15, 1967 through May 17, 1969

May 17, 1969

Contract Monitor: Caldwell P. Smith
Data Sciences Laboratory

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

Prepared
for

AIR FORCE CAMBRIDGE RESEARCH LABORATORIES
OFFICE OF AEROSPACE RESEARCH
UNITED STATES AIR FORCE
BEDFORD, MASSACHUSETTS 01730

ABSTRACT

Research into and the development of instrumentation for the investigation of factors affecting the quality of vocoded speech are documented. The work reported was specifically concerned with developing a better understanding of the role of the vocal source in the production both of synthetic speech and of natural speech. The design of and operating instructions for the VOTIF vocal track inverse filter - built as part of the program - are presented. A theoretical determination of the interaction between the vocal source and vocoder channel filters has been made and the effect of spectrum flattening on the peak factor and power of a vocoder channel have been computed. Lastly, the pulsed excitation of resonances is discussed. A form of pitch jitter which could either maximize vocal output or minimize vocal tract impedance effects is reported on.

FOREWORD

This report describes research and instrumentation development activities undertaken by SIGNATRON, Inc. of Lexington, Massachusetts to investigate factors in both natural and synthetic speech which could influence the quality of vocoded speech. These activities were carried out under Contract No. F19628-67-C-0292, beginning April 15, 1967 and ending May 7, 1969. The monitor of the contract was Mr. Caldwell P. Smith, CRBS, Air Force Cambridge Research Laboratories at Bedford, Massachusetts. Dr. Thomas H. Crystal of SIGNATRON was project director and principal investigator.

Many people other than the author of this report contributed to this program. Charles L. Jackson and Yogindiran Amarasingham participated in the assembly and testing of the VOTIF vocal track inverse filter. Donald S. Arnstein participated in the calculation of the effects of pitch jitter. The staff of Design Automation of Lexington, Massachusetts (through a subcontract) designed and constructed the VOTIF filtering units to SIGNATRON specifications. They also prepared the appendix to this report in which the design and operation of the filtering units is described.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

I. INTRODUCTION

This document reports on research and development done to investigate factors affecting the quality of vocoded speech. The work reported on was specifically concerned with developing a better understanding of the role of the vocal source in both the production of natural and the production of synthetic speech. The major part of the work was the development of instrumentation for performing experimental work in this area. Some theoretical investigations were also carried out.

1.1 VOTIF Instrumentation

The instrumentation developed has been designated as VOTIF for Vocal Tract Inverse Filter. VOTIF consists of a multi-unit analog filtering instrument and associated display and monitoring equipment. The filtering instrument is a cascade of units of two types. Null or anti-resonances are used to cancel vocal tract resonances or formants. A resonance is used to cancel the vocal tract anti-resonance introduced with an additional resonance, by coupling of the oral cavity with the nasal cavity.

VOTIF presently contains five operationally identical null units and one resonance unit. The frequencies and bandwidths of each unit are adjustable over the range shown in Figure 1.1. Both the frequency and the bandwidth of each unit may be set to a precision of within 0.5% of the frequency value. The readings obtained are within ±2% and ±10% of the actual frequency and bandwidth, respectively. Over a frequency from 100 Hz to 10 kHz, the transfer function is accurate to within ±0.25 dB of magnitude and ±0.10 milliseconds of delay. Full specifications and operating instructions for the filtering units are given in Appendix A of this report. These specifications, which were developed by SIGNATRON, are discussed in Section II. The display and monitoring equipment consists of a dual trace oscilloscope, a camera for the oscilloscope and a multi-function meter for checking signal levels, power supply levels and circuit resistance.

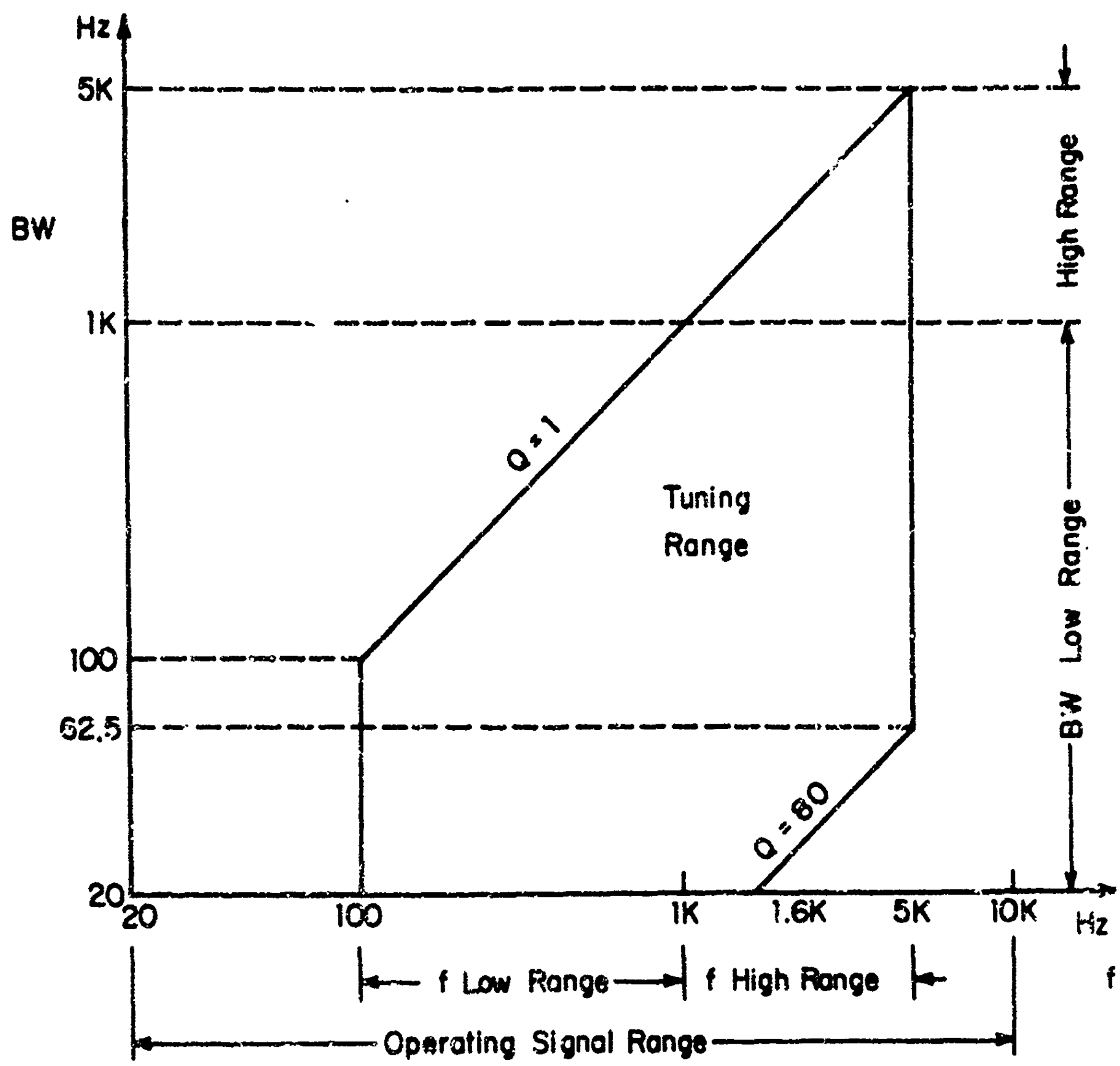


FIG. 1-1 TUNING RANGE OF FREQUENCY AND BANDWIDTH CONTROL SETTINGS

1.2 Theoretical Investigations

The theoretical researches done under this program all fall into the general area of Source-System Interaction. Such interaction exists both in the human and in synthetic speech systems. In synthetic speech systems it may exist in either or both the analyzer and the synthesizer. By source we refer to the vocal cords, in the human, or pitch generator, in a synthesizer (hiss excitation is not being considered). By system, we refer to the spectrum-shaping part of the production system. In the human, this is the vocal tract; in the synthesizer, the variable gain filters or the adjustable resonators. For convenience we will assume that the effect of glottal pulse shape is part of the system.

Previous consideration of source-system interaction has led to the improvement of channel vocoder speech through spectrum flattening, to debates on the origin of the residual ripple in inverse filtered speech, and to theoretical consideration of vocal source frequency optimized according to the tuning of the vocal tract (House, 1959). This program's consideration of source-system interaction was made in two areas. First, we considered source-system interaction in the channel vocoder. Secondly, we considered the excitation of resonators by periodic pulses.

1.2.1 Source-System Interaction in Channel Vocoders

Source system interaction in the channel vocoder results because the energy in any one of the analysis or synthesis bands is a function of the pitch frequency and pulse shape as well as the transfer function of the vocal tract. According to standard vocoder design techniques, this interaction is accepted in the analysis and compensated for in the synthesis by spectrum flattening. This procedure appears to work very well but is open to some questioning on theoretical grounds. The results of our investigations indicate that this compensation procedure should not generally be criticized because the order of the measured errors appear sufficiently low. Nevertheless we feel the questions discussed below were worth asking.

The first question relates to the digital encoding of the measured channel outputs of the analyzer. This encoding involves quantization of the analog measurements and in more sophisticated systems such as pattern matching vocoders — statistical reduction on the patterns. The question thus arises as to whether the quantization of spectrum information, as affected by pitch rate information which is also transmitted independently, seriously degrades the digital specification of the system information. In other words, would the quantization and transmission benefit from removal of the pitch rate information. For the pattern matching vocoder, we might also inquire if the pitch rate information, which is superimposed on the system information, appreciably increases the number of patterns which must be processed. In an attempt to clarify the question, the first part of Section 3 presents a determination of the amount of interaction. In terms of the 4 dB quantization steps commonly used in vocoder measurements, the effect appears not to be too serious, but such a determination is more properly made from actual trials rather than the theoretical considerations presented here. A doubt about this conclusion persists because, if the pitch rate were actually to have no effect on the analysis, spectrum flattening would not be needed at the synthesizer.

The second question raised pertains to the effect, on the synthesized speech waveform, of the spectrum flattening method commonly used. This method is the infinite clipping of the source signal after it has been filtered by one of a pair of channel filters for the channel. From theoretical considerations, it will be shown that this approach, in the worst case calculated, corrects the spectrum to within 2.5 dB of the desired power level. This is the expected effect of spectrum flattening. Less appreciated is the fact that spectrum flattening does not seriously distort the peak-factor of the signal. As will be shown below, the worst case calculated displays a peak-factor error of less than 2 dB.

1.2.2 Pulsing of Resonators

As noted above, a second consideration in the area of source-system interaction is that of pulsed excitation of resonances. There

is interaction in the sense that the amplitude of the resonator output can be optimized by proper selection of the pulse rate so that harmonics fall at the maximum of the resonance tuning curve. This phenomenon may be observed not only in the frequency domain but by calculations based on rotating vectors. We present these methods in Section 4.

An interesting extension of the above theory and observations gives a possible explanation of alternate period jitter in pitch periods. This phenomenon of alternately long and short pitch periods has been observed by Lieberman (1961) to occur in about 40% of vocalizations and has also been noted by Smith (1968) in selected data. As is explained in Section 4, the very occurrence of alternate period jitter doubles the number of spectral components, thus increasing the chance that a component will fall on or near the peak of the resonance tuning curve. The amount of the jitter can then be used to accentuate the specific component nearest the peak. That this is the controling factor in actual pitch jitter is a matter of hypothesis. The theory, however, leads to formulas for the calculation of jitter as a function of pitch and formant frequency and thus provides a basis for subsequent verification.

The topic of vocal energy optimization bears some discussion. The suggestion that this may actually occur implies the existence, as part of the human speech production system, of a measurement and control mechanism for sensing and improving vocal efficiency. While this may seem improbable on a neurological basis it could occur on a physical basis. Physical systems tend to operate in modes which minimize certain types of energy. As a coupled physical system, the larynx and the vocal tract could function in this matter. On the other hand, the maxima of vocal tract transmission are also maxima of vocal tract impedance. The result is a tendency of the vocal tract to resist being driven at rates producing components falling on the resonances (Crystal, 1966). Simple modification of the jitter formulas can lead to determination of amounts of jitter which reduce a component which would otherwise occur at a resonance.

A third facet of the program described by this final report was the intended computer simulation of a model of the Vocal Response Synthesizer (VRS) vocoder synthesizer. This facet of the program was discontinued when it appeared more advantageous to devote program resources to the other areas.

II. INVERSE FILTERING WITH VOTIF

2.1 Background

The concept of inverse filtering is a natural consequence of the acoustic theory of speech production (Fant, 1960). The theory of production describes the vocal tract as a mechanism for performing linear, minimum phase, acoustic filtering of the air flow through the glottis. The filter is characterized by having an infinite number of poles or resonances located, on the average, at the odd harmonics of 500 Hz. In general, during vocalization, only the first three or four of these resonances are excited, with an extra pole and stable zero (anti-resonance) entering into the filter during the production of nasal sounds. A natural consequence of this theory is that each significant pole may be canceled with a zero (or anti-reson or null) of the same frequency and bandwidth. Likewise, the zero may be cancelled by a pole. One verification and application of acoustic theory of speech production is the successful construction and use of inverse filters by other researchers [Mathews, et.al. (1961), Holmes (1962) and Linqvist (1964 and 1965)].

VOTIF was built to provide the Digital Speech Branch of AFCRL with the equipment to study vocal source characteristics for their possible effect on vocoded speech quality. In building this equipment we sought to utilize the latest in solid state technology, be able to handle wide-band speech, permit the use of direct-reading linear controls and give ease of calibration. The specific design considerations, circuitry and operating instructions for the filters appear as Appendix A to this report.

2.2 Design Considerations

2.2.1 Performance Specifications

The target specifications, which were often exceeded in the instrument itself, were derived from considerations of both the human speech production and hearing mechanisms as previously characterized

by other researchers.

1. Tuning Range

Tuning range is presented in Figure 1.1. The lower bound on the frequency is one cited by Flanagan (1965) as a design criterion for formant vocoders and is a little over half the lowest formant frequency ($\approx$190 Hz) measured by Peterson and Barney (1952). The upper limit of the tuning range will permit matches to most fourth formants and provide for a sharp glottal pulse.

2. Precision and Accuracy

The criterion for choosing the precision is that the adjusted values of frequency and bandwidth must approach the target values closely enough so that the ripple remaining from incomplete cancellation will not seriously distort the waveform of the glottal pulse. In this case the ripple was evaluated by looking at the area under the maximum lobe of the ripple and saying that this area should not exceed 2.5% of the area of the desired response. This ripple is obtained by first finding the Laplace transform of the combined transmission of resonance and null

$$G(s) = H(s) \cdot P(s) = \frac{(s+b+\varepsilon)^2 + (a+\delta)^2}{(s+b)^2 + a^2}$$

$$= 1 + \frac{2\varepsilon(s+b) + 2\delta \cdot a}{(s+b)^2 + a^2} + \frac{\varepsilon^2 + \delta^2}{(s+b)^2 + a^2}$$

where

ε = error in adjusting bandwidth (radian)

δ = error in adjusting frequency (radian)

For small ε and δ the last term is negligible and we have for the impulse response

$$g(t) = u_o(t) + e^{-bt}\{2\varepsilon \cdot \cos at + 2\delta \cdot \sin at\}$$

Looking at just one lobe of the sinewave, we see that the area under it is $4\delta/a$. For $\varepsilon_{max} = \delta_{max}$ the maximum area under a lobe is $5.78/a$. Allowing a maximum allowable one-lobe area of .025 (the impulse has unit area), we get the relationship

$$\frac{5.7\delta_{max}}{a} \leq .025$$

or

$$\epsilon_{max} = \delta_{max} \cong .005\ a$$

From this it can be seen that the required frequency precision is 1/2% of measured value.

Accuracy requirements reflect how closely we wish to know the true parameters for the resonance. Suitable criteria appear to be the DL's for formant frequencies and bandwidths as reported by Flanagan (1965, pp. 212-213) in discussing his own (Flanagan, 1955) and Stevens' (1952) experiments. Frequency DL's of 3 to 5 percent and bandwidth DL's of 20 to 40 percent are just discriminable.

3. Operating Range and Characteristics

The maximum frequency of 10 kc was chosen so that there would be ample resolution for extracting timing information from the glottal signal. Lieberman (1961) has noted interesting laryngeal behaviour which produces timing shifts in the glottal pulse of the order of tens of milliseconds.

The lower bound is chosen such that there will be a stable base-line over several pitch periods yet the complexities of going to DC operation will be avoided.

The delay criteria was chosen so as to preserve timing information as discussed above.

The amplitude criteria was chosen so that observed amplitudes in unsupressed components, such as the one due to larynx-vocal tract interaction, will be accurate to approximately 3%.

4. Gain

In modeling the vocal tract as an acoustic system, one notes that its transmission at DC is unity. Thus, its inverse should also have the capability of being adjusted to unity transmission at DC.

5. Signal-to-Noise Ratio

Chosen to match performance characteristics of other audio equipment and be reasonable in terms of the technology utilized.

2.2.2 Other Design Considerations

An important consideration in the design of VOTIF was the use of resistive controls. In the present circuitry this gives the precision and accuracy of adjustment desired and allows for adjustment and calibration by appropriate resistive trimming. The use of resistors also has implications for extending the capability of VOTIF. One extension is to provide for automatic recording of the frequency and bandwidth settings. This can be achieved either by momentary switching of an adjustment resistor from the filtering circuit to a measuring circuit or by adding a third gang to each pot for continuous connection to the measuring circuit. For automatic adjustment of the filtering circuits, the potentiometers could be replaced by digital attenuators. These attenuators are merely D-to-A converters in which the constant reference source has been replaced by the signal to be attenuated.

A design objective which was rejected after careful consideration was the implementation of units that could be switched between null and resonance behavior. Considered for implementation was the use of one type of circuit either directly or in a feedback loop, to get its inverse. The strict constraints on phase over the wide bandwidth of the instrumentation obviated this approach. Hence, two separate types of units were designed and built.

2.3 Use of VOTIF

2.3.1 Planned Use on Speech

The use of VOTIF on natural speech requires the implementation of a distortion-free means for repeating short segments of the signal to be analyzed. The segments should be several pitch periods in length so that any initial transients may die out. However, the segments should be short enough so that the repetition rate is adequate. An adequate repetition rate will permit close coordination of filter adjustment and observation of the effect

of the adjustment. One would also like to avoid flicker but this is not generally obtainable with low pitch signals. Besides reproducing the speech signal, the repetition instrumentation should provide signals for jitter-free triggering of the display. Two means for implementing the desired signal reproducing instrumentation are discussed in the following. Neither was implemented nor tested as part of the work performed. Rather, VOTIF was tested with synthetic signals.

Previous applications of the inverse filters have utilized FM tape reproducers for repetitive presentation of the signal to be analyzed (Lindquist, 1964). FM is used where AM cannot be, because the FM techniques preserve waveform whereas AM techniques introduce appreciable phase distortion in order to preserve relatively flat amplitude vs frequency characteristics. Tape recording techniques do however possess the drawback that the mechanical design requires tape loops of lengths which keep the repitition rate low. In addition, there would be problems indexing through long signals so as to give an analysis of many consecutive periods of a long vocalization. There also is a question of the stability of the recording tape and the reproduced signal from period to period.

An alternative approach is to use a digitally stored representation of the signal to be analyzed. Repetitive D-to-A conversion is performed to obtain the analog signal for analysis. When the digital signal has been obtained directly or from an FM recording, the requirement for a phase-distortion-free signal is met. Long utterances recorded on digital tape or disk may be easily indexed to provide continuous analysis and the actual segment length repeated can be chosen to optimize the analysis. At a 10 kHz sampling rate, only 1000 storage locations are needed to provide a tenth of a second segment, which would provide at least two full pitch periods of a pitch having as low a frequency as 50 Hz. With the present general availability of digital hardware, this approach is highly advisable.

2.3.2 Use of VOTIF on Synthetic Signals

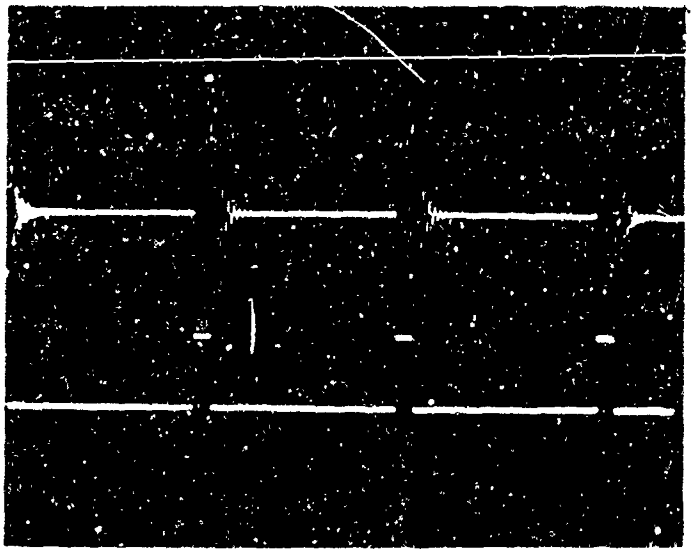
To demonstrate the use of VOTIF in processing signals, two types of experiments were run. In the first, the cascade of a VOTIF resonance and a VOTIF null were excited by a pulse generator, to demonstrate the inverse characteristics of these two types of networks. In the second experiment, a synthetic two-formant vowel was analyzed.

The results of the experiment with the paired VOTIF resonance and VOTIF anti-resonance are illustrated in the three photographs of Figure 2.1. These photographs show the VOTIF input and output for three different conditions. In all pictures the bottom oscilloscope trace is the pulse generator input signal to the system; the top, the processed signal. The pulses come from a General Radio Model 1340 generator and are 1 msec wide and occur at a rate of 100 pps.

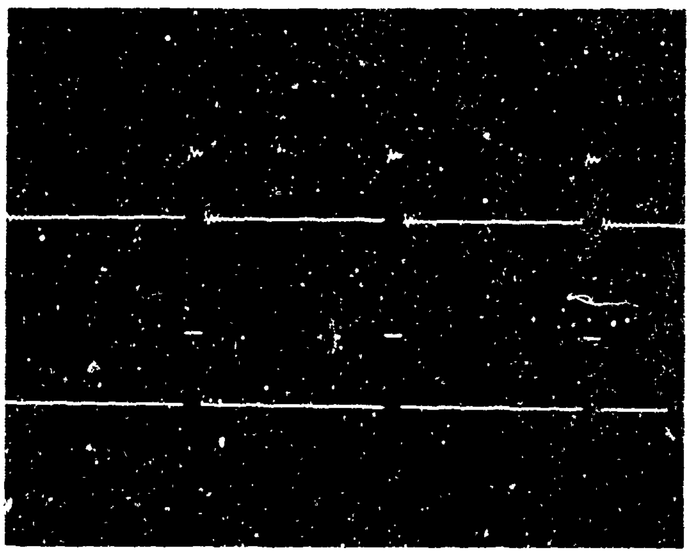
In the top photograph (Fig. 2,1a) only the resonance unit is in the circuit. It is set for a frequency of 3600 Hz and a bandwidth of 665 Hz. In Fig. 2.1b, the null has been switched into the cascade following the resonance. The null is set to F = 3350 and BW = 675, giving only partial cancellation due to the frequency mistuning of 7%.

In Fig. 2.1c, the resonance has been totally cancelled with the null set to F = 3650 and BW = 675. The null settings differ from the resonance settings by about 1.5% in both frequency and bandwidth. This is well within the design specifications. There is slight overshoot at the edges of the pulse due to incomplete cancellation for the very large derivatives occurring at these edges. The system noise tends to widen the oscilloscope trace.

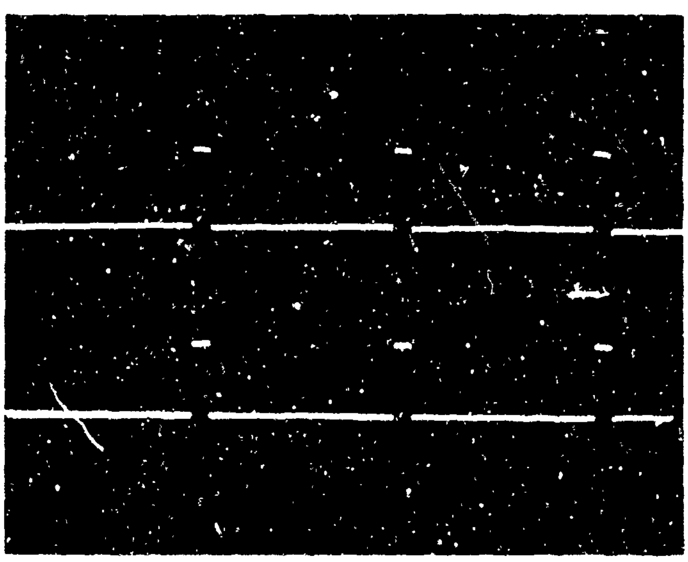
In the second experiment, a two formant synthetic vowel sound was analyzed using null units only. The signal was generated by a Bell System Science Experiment No. 3, speech synthesizer and the above-referenced pulse generator.



a) Uncancelled resonance of
F = 3600 Hz
BW = 665 Hz



b) Partially cancelled resonance with null of
F = 3350 Hz
BW = 675



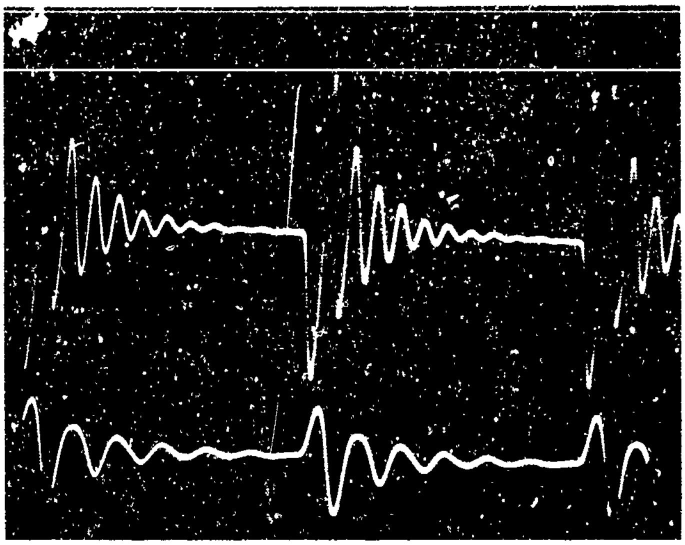
c) Cancelled resonance with null of
F = 3650 Hz
BW = 675 Hz

Fig. 2.1
Cancellation of VOTIF Resonance by VOTIF Null with 1 msec, 100 pps pulse input. Bottom trace of all pictures shows pulses.

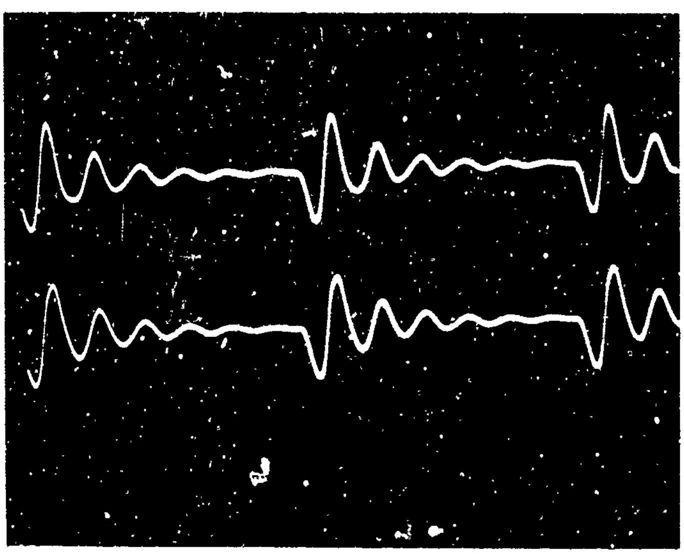
The synthesizer utilizes RLC tuned circuits to simulate the formants. An external pulse generator was used for the periodic source. Low pass filters were used at both the input and at the output of the cascaded nulls, to help reduce noise. The results of the experiment are illustrated in the three photographs of Figure 2.2. In all pictures, the bottom trace is the unprocessed signal. The repetition rate is 100 pps.

In Fig. 2.2a we show the effect of removing the first formant at F = 695 and BW = 150. What remains is the damped exponential for the second formant. In Fig. 2.2b, we show the effect of cancelling the second formant at F = 1440 and BW = 740. What remains in this case is the first formant. The similarity of the second formant to the unprocessed signal indicates the weakness of the second formant produced by the synthesizer.

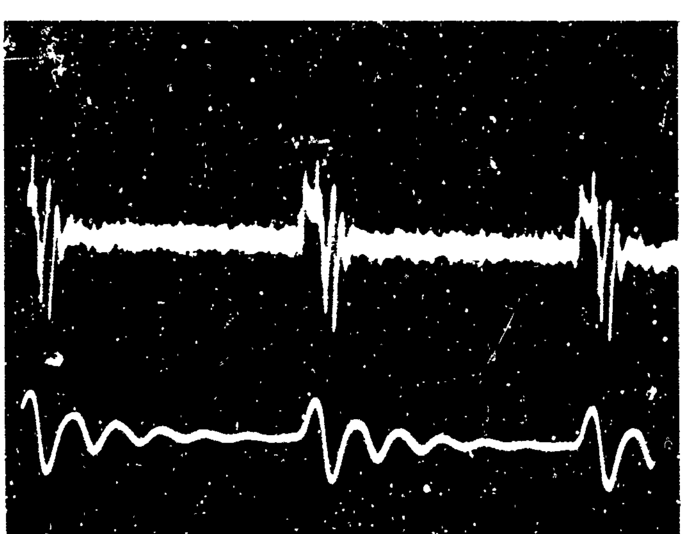
Figure 2.2c illustrates the cancellation of both formants. The resulting pulse represents the original source pulse as modified by the amplifiers and low-pass filters. Unlike natural speech, the synthetic source is a sharp-edge pulse of short duration and when rederived by inverse filtering exhibits spike type overshoot as discussed in the previous experiment. The noise in the inverse filtered signal is high frequency synthesizer and amplifier noise, amplified by the rising gain-frequency characteristics of the nulls. The noise may appear to be sinusoidal because of a transfer function peak around 18 kHz caused by the intersection of the rising null gain with the 18 kHz low-pass filter in the null output stage. It should be noted that in adjusting the filter units it is important not to overload the internal circuits of each filter. The test points described in the appendix are particularly useful for monitoring for overload.



a) Signal after
cancellation of
first formant:
F = 695 Hz
BW = 150 Hz



b) Signal after
cancellation of
second formant
F = 1440 Hz
BW = 750 Hz



c) Signal after
cancellation
of both
formants

Fig. 2.2
VOTIF analysis of two-formant synthetic speech. Bottom trace of all pictures shows synthetic vowel.

III. SOURCE SYSTEM INTERACTION IN THE CHANNEL VOCODER

Source-system interaction in the channel vocoder is the effect of the repetition rate of the source on the output of the channel filters. As there are channel filters in both the analyzer and the synthesizer portion of the vocoder it may occur in both. In the analyzer the interaction would be that between the human vocal source and the analyzing filters. In the synthesizer, it is that between the synthesizer buzz source and the synthesis filters.

If this interaction were to take place in both the analyzer and the synthesizer it would distort the spectrum of the synethesized speech. It must, therefore, be compensated during either analysis or synthesis. In presently used vocoder techniques, it is compensated in the synthesizer by spectrum flattening. This means that the channel signals transmitted from analyzer to synthesizer carry some unnecessary information about the pitch rate. To give an indication of the amount of the source-system interaction component in the channel signals and the needed amount of correction at the synthesizer, the following section presents a calculation of this component. The section after next discusses the effect of spectrum flattening on the resulting synthesized signal in terms of both the degree of normalization of power and the modification of signal peak factor.

3.1 The Effect of Pitch Rate on Channel Filter Output

The effect of pitch rate on channel filter output is a function of the number of components passing through a particular filter and the expected number. For a pulse rate of ω_o radians/second we would expect a filter of Ω radians bandwidth to pass Ω/ω_o components, which is not necessarily integer. However, the actual number of components passed must be an integer and is given by

$$N = \left\lfloor \frac{\omega_\mu}{\omega_o} \right\rfloor - \left\lceil \frac{\omega_\ell}{\omega_o} \right\rceil \qquad (3.1)$$

SIGNATRON®

where ω_μ and ω_ℓ are the upper and lower limits of the passband, respectively. They are related by

$$\Omega = \omega_\mu - \omega_\ell \tag{3.2}$$

From these formulas we see that N is bounded as follows:

$$\left|N - \frac{\Omega}{\omega_o}\right| \lesssim 1 \tag{3.3}$$

The absolute difference between N and Ω/ω_o may actually approach arbitrarily close to 1.

If we consider that the interaction is the ratio of the actual signal power passed by the filter to the expected signal power and that each component adds one unit of power we would get

$$I_{dB} = 10 \log_{10} \left[N\frac{\omega_o}{\Omega}\right] \tag{3.4}$$

From Eq. (3.3) we get a bound on I

$$10 \log_{10} \left(\frac{N}{N+1}\right) \leq I_{db} \leq 10 \log_{10} \left(\frac{N}{N-1}\right) \tag{3.5}$$

The upper bound does not exist for N=1.

The possible range of I for various small values of N is given in Table 3.1. The values of N represented in the table are typical for the number of components that fall in the various channel bands in vocoders.

The interaction for N from 1 to 3 is of the order of the quantum step used in quantizing the vocoder analyser output. This indicates that different pitches could result in more than one pattern of digits for a given articulation of a particular speaker. The interaction may also be interpreted as the error which exists if spectrum flattening or some similar form of compensation is not used in a vocoder system. That this error is appreciable can be demonstrated

by the subjective improvements obtained by using spectrum flattening. Both effects of this type of interaction are increased by the dynamics of changing pitch. Thus changes of the order of the ranges listed in the table below would occur every time a pitch change caused a component to move from one band to an adjacent one.

Table 3-1

VARIATION OF FILTER OUTPUT INTENSITY FROM EXPECTED VALUE AS A FUNCTION OF THE NUMBER OF COMPONENTS PASSED

N	I_{min} (dB)	I_{max} (dB)	Range (dB)
1	-3.0	-	-
2	-1.8	3.0	4.8
3	-1.2	1.8	3.0
4	0.9	1.2	2.1

SIGNATRON®

3.2 The Effect of Spectrum Flattening on the Synthesized Signal

Spectrum flattening as performed in channel vocoder synthesizers is achieved by distorting the waveform of the signal. In analyzing spectrum flattening, one should investigate the effect of the flattening on the shape of waveform as well as on the power of the waveform. In the following, we examine peak factor -- the ratio of peak signal to signal power -- as an indicator of the effect on the waveform.

A model of a single channel of a vocoder is shown in Fig. 3.1. The two bandpass filters are identical, with the result that the same frequency components appear at both (A) and (C), but with their strengths changed. Due to the action of the infinite clipper many more components appear at (B). The power at (B) is one because the signal there is always either +1 or -1. Because some of the components contributing to this power do not pass through BPF_2, the power at (C) is actually lower than the target value of unity.

For a constant frequency impulse source and ideal bandpass filters the signal at A is

$$S_A(t) = \frac{\sin\left(N \frac{\omega_o t}{2}\right)}{\sin\left(\frac{\omega_o t}{2}\right)} \cos \omega_c t \qquad (3.6)$$

where N = number of components passed by the filter,

ω_o = radian pulsing frequency i.e., difference in frequency between adjacent components, and

ω_c = is the center frequency of the passed components.

When the number of components N is odd, ω_c is the frequency of the center component. When N is even, ω_c is the average of the two innermost components.

Because of the even symmetry, the peak-signal occurs for t = 0 and has a value N, which is actually the sum of the N equal amplitude

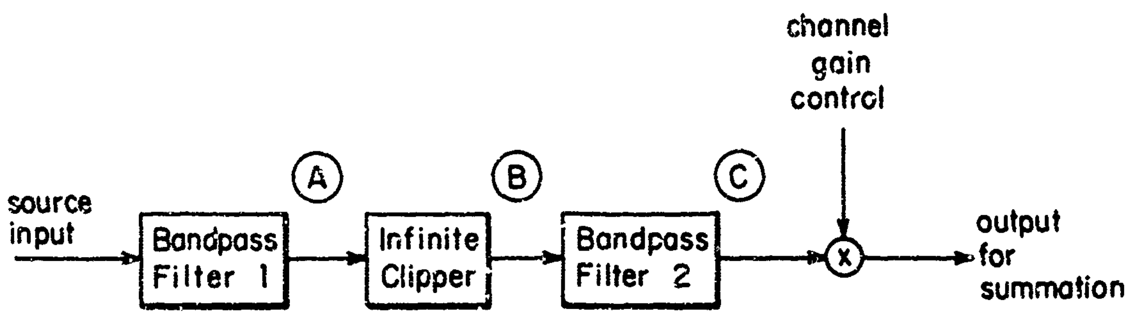


FIG. 3-1 MODEL OF SINGLE CHANNEL OF SPECTRUM FLATTENING SYNTHESIZER

components. The total power is N times the power in a single component, this power being normalized to one. Thus the peak factor, defined as

$$PF = 10 \log_{10} \left(\frac{\text{peak}^2}{\text{power}}\right) \tag{3.7}$$

is $10 \log_{10} N$.

Because the signal symmetry is maintained during the clipping operation and subsequent filtering, the peak signal value continues to be the sum of the individual component amplitudes. The power is the sum of the squares of these amplitudes, giving

$$PF = 10 \log_{10} \left[\frac{(\Sigma c_n)^2}{\Sigma c_n^2}\right] \tag{3.8}$$

where c_n is an individual component amplitude.

To obtain the value of the components we analyze sgn $[S_A(t)]$, as given by Eq. (3.6), at each of the components. We define sgn $\{\cdot\}$ as

$$\text{sgn}\ \{x\} = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x = 0 \\ -1 & \text{for } x < 0 \end{cases} \tag{3.9}$$

The strength of the component is

$$c_{\pm n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{\text{sgn}\left[\frac{\sin\left(\frac{N\theta}{2}\right)}{\sin\left(\frac{\theta}{2}\right)}\right] \cos(\hat{n}\theta)\right\} \{\cos(\rho\theta)\ \text{sgn}[\cos(\rho\theta)]\}\ d\theta$$

$$\mp \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{\text{sgn}\left[\frac{\sin\left(\frac{N\theta}{2}\right)}{\sin\left(\frac{\theta}{2}\right)}\right] \sin(\hat{n}\theta)\right\} \{\sin(\rho\theta)\ \text{sgn}[\cos(\rho\theta)]\}\ d\theta \tag{3.10}$$

where $\theta = \omega_o t$

$$\rho = \frac{\omega_c}{\omega_o} \geq 1$$

$$\hat{n} = \begin{cases} n \text{ for } N \text{ odd} \\ \frac{1+2n}{2} \text{ for } N \text{ even} \end{cases}$$

The magnitude of the index of c indicates the distance of the component being evaluated from ω_c, the center frequency of the components. The sign indicates whether the component is lower or higher than the center.

For $\rho >> N$, we can replace the second terms in each integral of Eq. (3.10) by their averages which are $\frac{2}{\pi}$ and 0, for the first and second integrals, respectively. Thus, for large ρ, the strength of components equidistant from ω_c would be equal. This is to be expected because letting $\rho >> N$ is equivalent to saying that the center frequency of the passband is much higher than its bandwidth and non-linear distortion does not cause interaction between symmetrical components.

The evaluation of the integrals of (3.10) is accomplished by piece-wise summing integrals of that portion of the argument where the sgn $\{\cdot\}$ functions in the integral do not change sign. Because of symmetry, it is necessary to integrate only from 0 to π. This allows the reduction

$$\operatorname{sgn}\left[\frac{\sin\left(\frac{N\theta}{2}\right)}{\sin\frac{\theta}{2}}\right] \rightarrow \operatorname{sgn}\left[\sin\left(\frac{N\theta}{2}\right)\right] \tag{3.11}$$

Thus Eq. (3.10) reduces to

$$c_n = \frac{1}{\pi}\sum_i \mathrm{SGN}(\theta_i^+)\int_{\theta_i}^{\theta_{i+1}} \cos(\hat{n}\theta)\cos(\rho\theta)\,d\theta$$
$$\mp\frac{1}{\pi}\sum_i \mathrm{SGN}(\theta_i^+)\int_{\theta_i}^{\theta_{i+1}} \sin(\hat{n}\theta)\sin(\rho\theta)\,d\theta \tag{3.12}$$

where

$$\mathrm{SGN}(\theta) = \mathrm{sgn}\left[\sin \frac{N\theta}{2}\right] \mathrm{sgn}\,[\cos(\rho\theta)]$$

and where the θ_i's define points of change of SGN(θ) for $0 \leq \theta \leq \pi$.

The integrals in Eq. (3.12) have the values

$$\int \cos(\hat{n}\theta)\cos(\rho\theta)\,d\theta = \begin{cases} \dfrac{\sin[(\rho - \hat{n})\theta]}{2(\rho - \hat{n})} + \dfrac{\sin[(\rho + \hat{n})\theta]}{2(\rho + \hat{n})}, & \rho = \hat{n} \\[2ex] \dfrac{\theta}{2} + \dfrac{\sin(2\rho\theta)}{2\rho}, & \rho = \hat{n} \end{cases} \tag{3.13a}$$

$$\int \sin(\hat{n}\theta)\sin(\rho\theta)\,d\theta = \begin{cases} \dfrac{\sin[(\rho = \hat{n})\theta]}{2(\rho - \hat{n})} - \dfrac{\sin[(\rho + \hat{n})\theta]}{2(\rho + \hat{n})}, & \rho = \hat{n} \\[2ex] \dfrac{\theta}{2} - \dfrac{\sin(2\rho\theta)}{2\rho}, & \rho = \hat{n} \end{cases} \tag{3.13b}$$

Thus the evaluation of the component strengths can be reduced to a summation which can be performed on a computer. The computer can also be programmed to determine the θ_i's.

We now consider calculation of the limiting case of $\rho \to \infty$ to derive formulas which not only give us additional feeling for the mathematics but also provide a means for checking calculations performed according to the above equations. As above, we integrate from 0 to π and reduce the second term of the integral to the constant $\frac{2}{\pi}$. This gives

$$c_n = \frac{2}{\pi^2}\int_{0+}^{\pi} \mathrm{sgn}\left[\sin\left(\frac{N\theta}{2}\right)\right]\cos(\hat{n}\theta)\,d\theta \tag{3.14}$$

This is further reduced to a summation by the piece-wise integration methods described above. This gives

$$c_n = \begin{cases} \frac{2}{\pi} & \text{for } N = 1 \\ \frac{2}{\pi^2} \sum_{k=0}^{\left[\frac{N}{2}\right]-1} (-1)^k \int_{\frac{2k\pi}{N}}^{\frac{2(k+1)\pi}{N}} \cos(\hat{n}\theta)\, d\theta + (-1)^{\frac{N}{2}} \int_{\frac{2}{N}\left[\frac{N}{2}\right]\pi}^{\pi} \cos(\hat{n}\theta)\, d\theta & \text{for } N \geq 2 \end{cases} \tag{3.15}$$

where $[x]$ = integer value of x.

The integrals may be reduced using

$$\int_{\frac{2k\pi}{N}}^{\frac{2(k+1)N}{N}} \cos(\hat{n}\theta)\, d\theta = \begin{cases} \frac{2\pi}{N} \text{ for } \hat{n} = 0 \\ \frac{1}{\hat{n}} \left\{ \sin\left[\frac{2\hat{n}(k+1)\pi}{N}\right] - \sin\left[\frac{2\hat{n}k\pi}{N}\right] \right\} & \text{for } \hat{n} \neq 0 \end{cases} \tag{3.16}$$

Which leads to an easily implemented computational procedure.

The results of the computations outlined above are shown in Figs. (3.2 and 3.3). In Fig. (3.2) is shown the power in the components after spectrum flattening, for various values of ρ. We can see that the spectrum flattening achieves its objective to within 2.5 dB. As noted above, the power output is less than 0 dB because the bandpass filter after the clipper removes some of the components which contributed to the 0 dB power level at the output of the clipper.

In Fig. (3.3), the peak factor of the channel output signals is shown. In this case the computed peak factor is within 2 dB of the peak factor obtained without clipping. The conclusion to be drawn is that spectrum flattening, as modeled above, is an effective way of dealing with source-system interaction in channel vocoders. This acceptance is conditioned on there being no source-system interaction distortion in the encoding process, as discussed above.

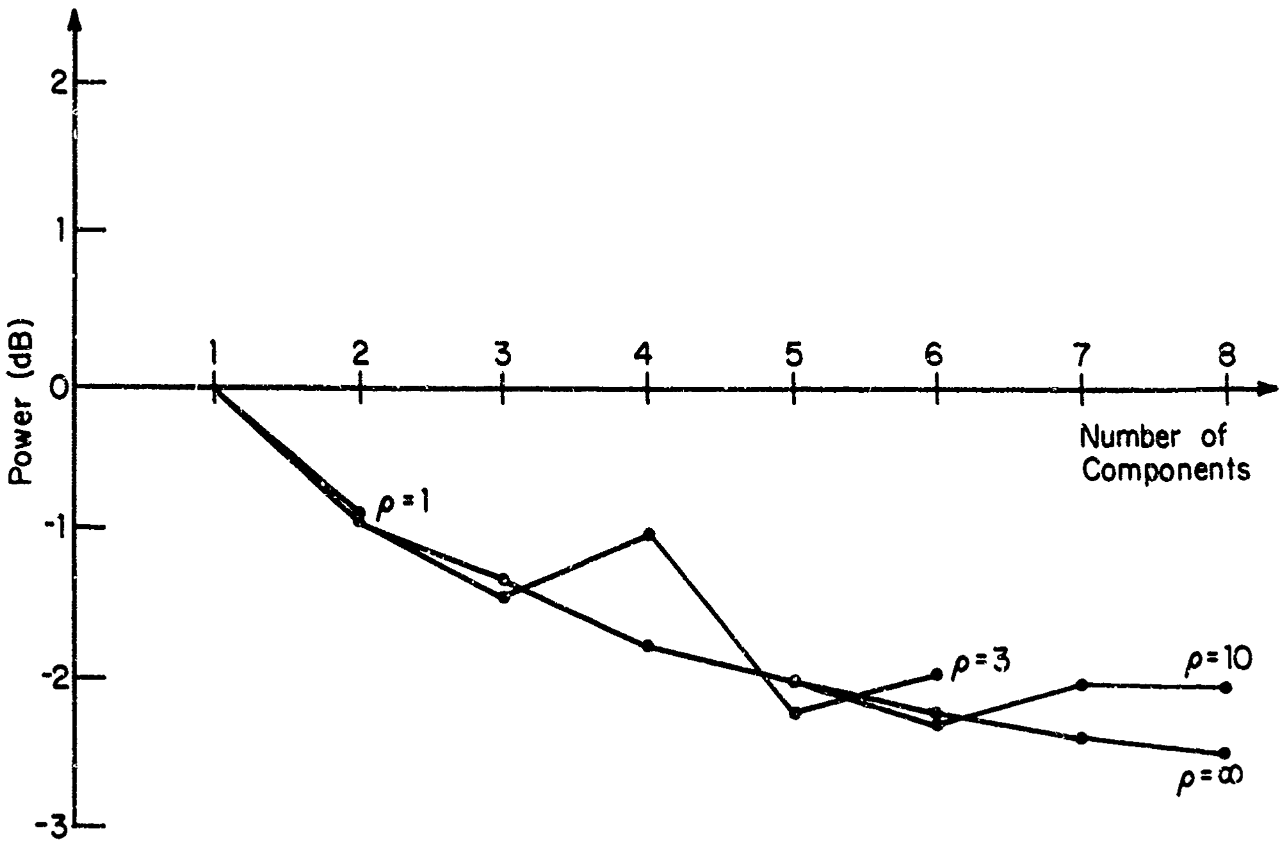


FIG. 3-2 EFFECT OF SPECTRUM FLATTENING ON CHANNEL POWER.
ρ is the ratio of center frequency to fundamental frequency.

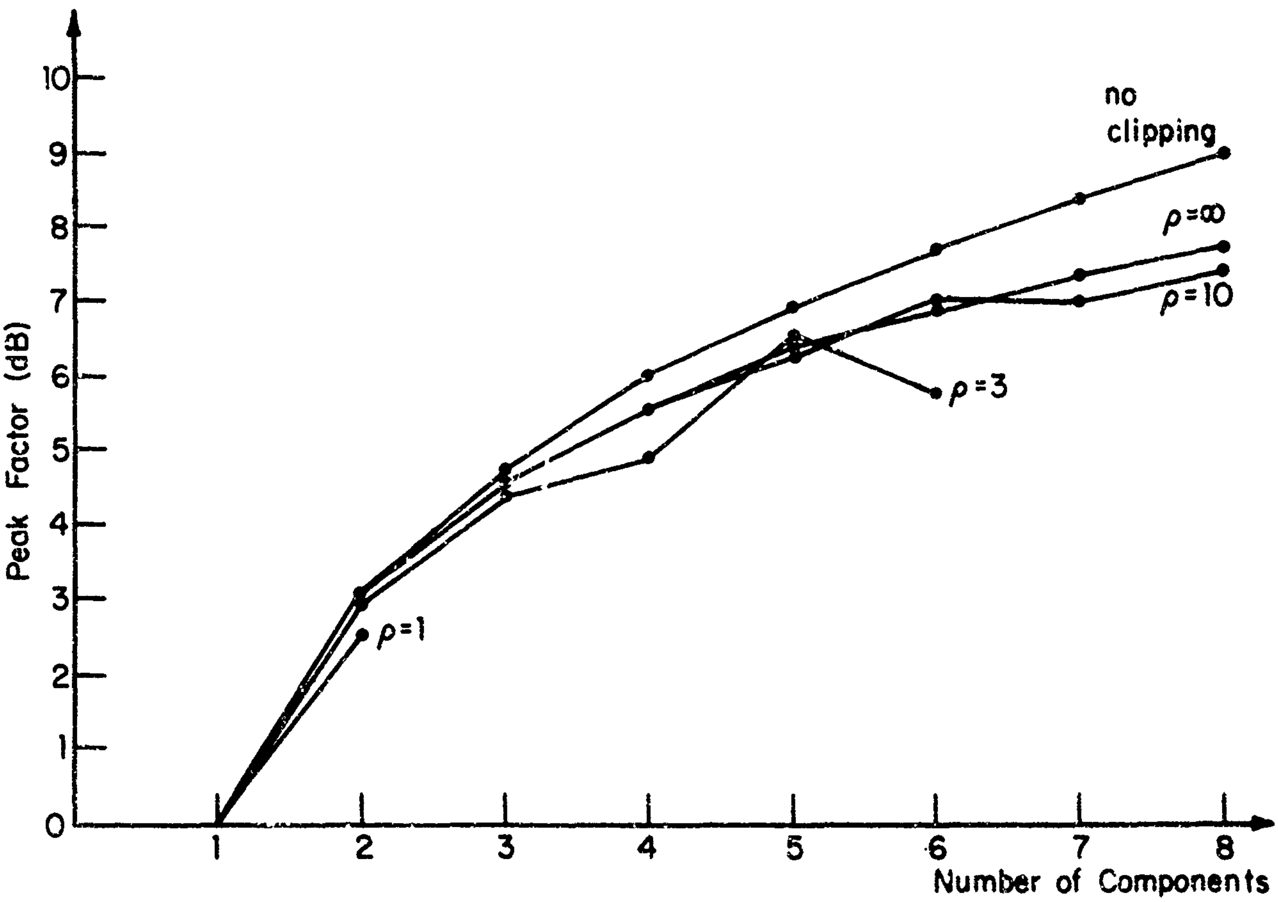


FIG. 3-3 EFFECT OF SPECTRUM FLATTENING ON PEAK FACTOR. ρ is the ratio of center frequency to fundamental frequency.

IV. PULSING OF RESONATORS

Our interest in the periodic or quasi-periodic impulsing of a harmonic oscillator or resonator derives from its similarity to the vowel production process. For most vowels, the first formant dominates the generated signal. Hence, we may hope to obtain interesting results from the study of a single oscillator. In actual speech production the oscillation appears to derive its excitation from a single discontinuity in the glottal pulse. This discontinuity can be replaced by an impulse if the resulting amplitude is scaled by the appropriate power of the frequency of oscillation and the phase is shifted by a multiple of $\pi/2$ radians. The power to which the frequency is raised and the multiplier of the phase shift is equal to the order of the discontinuity. In the discussion which follows, this compensation of amplitude and phase is unimportant.

In what follows, we examine how the amplitude of the oscillation varies as a function of the relationship between the pulse rate and the oscillator frequency. In a second section, we explain how appropriate alternation of short and long inter-pulse periods may moderate maxima or minima of resonator response.

4.1 Periodic Pulsing of a Resonator

As was described in a paper by House (1959) changing pulse rate, while holding the resonance characteristics constant, produces fluctuations in the amplitude of the signal transmitted through the resonance. This can be explained by Fig. 4.1 in which we show how the transmission function of a resonance effects the amplitude of the components of impulse trains of two different frequencies as shown by solid and dashed lines respectively. The pulse rate represented by the dashed line will produce a larger output than the other because a component falls at the peak of the transmission.

Another way of examining this phenomenon is in terms of the complex representation for the behavior of the harmonic oscillator between pulses:

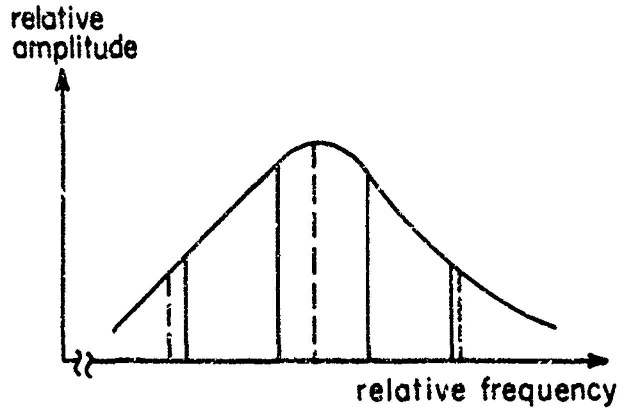


FIG. 4-1 TRANSMISSION OF COMPONENTS BY A RESONANCE

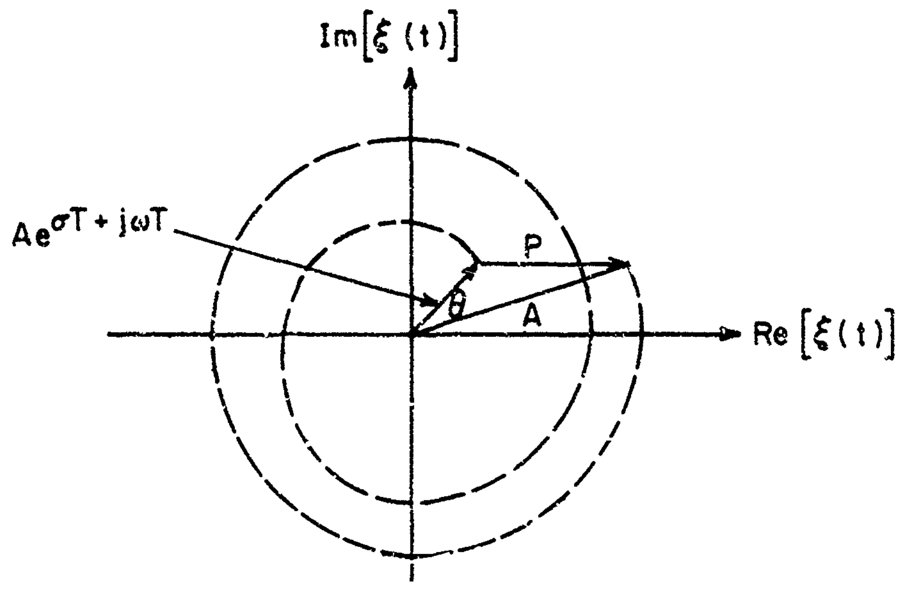


FIG. 4-2 HARMONIC OSCILLATOR BEHAVIOR

$$\xi(t) = Ae^{(\sigma + j\omega)c} \tag{4.1}$$

where A is a complex amplitude. The real signal which would actually be obtained from a resonator is the real part of this complex signal. If we pulse the oscillator every T seconds with a real valued pulse of amplitude p, the steady-state oscillator behavior may be described by the equation

$$A = Ae^{\sigma T + j\omega T} + p \tag{4.2}$$

This equation indicates that the ringing of the oscillator starts at a value A and rings for T seconds until it achieves a value $A\exp[\sigma T + j\omega T]$. At such time a pulse p is used to re-obtain the initial oscillator amplitude and a new period of decay begins.

Equation (4.2) is illustrated geometrically in Fig. 4.2. The spiral shows the locus of $\xi(t)$ over the interval T. The real signal is the real axis projection of the vector whose tip follows the spiral. The rotation is the angular change of the sinusoid while the decreasing diameter of the spiral is the exponential decay of the amplitude of the sinusoid. The angle θ between the two vectors is the total rotational angle modulo 2π.

The response or ratio of the oscillator amplitude to the pulse amplitude may be obtained from Eq. (4.2):

$$\frac{A}{p} = \frac{1}{1 - e^{\sigma T + j\omega T}} \tag{4.3}$$

From this we may obtain the squared magnitude of the response.

$$\left|\frac{A}{p}\right|^2 = \frac{1}{1 - 2e^{\sigma T}\cos\omega T + e^{2\sigma T}} \tag{4.4}$$

Note that this equation could also be obtained by applying trigonometry to the vector diagram in Fig. 4.2.

From Eq. (4.4) it can be seen that the magnitude of the response oscillates between maxima and minima as ωT changes through successive multiples of π. We have minima

$$\left|\frac{A}{P}\right|^2 = \frac{1}{(1 + e^{\sigma T})^2} \tag{4.5}$$

for $\omega T = (2n+1)\pi$

and maxima

$$\left|\frac{A}{P}\right|^2 = \frac{1}{(1 - e^{\sigma T})^2} \tag{4.6}$$

for $\omega T = 2n\pi$

This alternating maximization and minimization of the response is the same as that predicted by our previous discussion of frequency components and calculated in detail by House (1959).

A set of curves depicting Eq. (4.4) is given in Fig. 4.3. In labeling these curves we have used the relationships

$$2\pi FT = \omega T$$

$$BW \cdot T = \frac{\sigma}{\pi} T$$

where F and BW are the frequency and bandwidth of the resonator, respectively, and T is the period of the pulses. The amplitude of the response in dB is shown on the vertical axis and the normalized quantity $BW \cdot T$ on the horizontal axis. The functional relationship between these two quantities is shown for six values of $2\pi FT$, the argument of the cosine in Eq. (4.4). At BW = .5 there is a change of vertical scale.

The open circles and dashed line on the graph illustrate how it is used to obtain the response for a fixed resonator as the pulse rate is varied. (Pulse rate is the reciprocal of T.) The illustration is for F = 300 and BW = 50. Each circle represents a different frequency. Scanning from left to right the maximum of response occur at 300 Hz, 150 Hz, and 100 Hz; the minima at 200 Hz

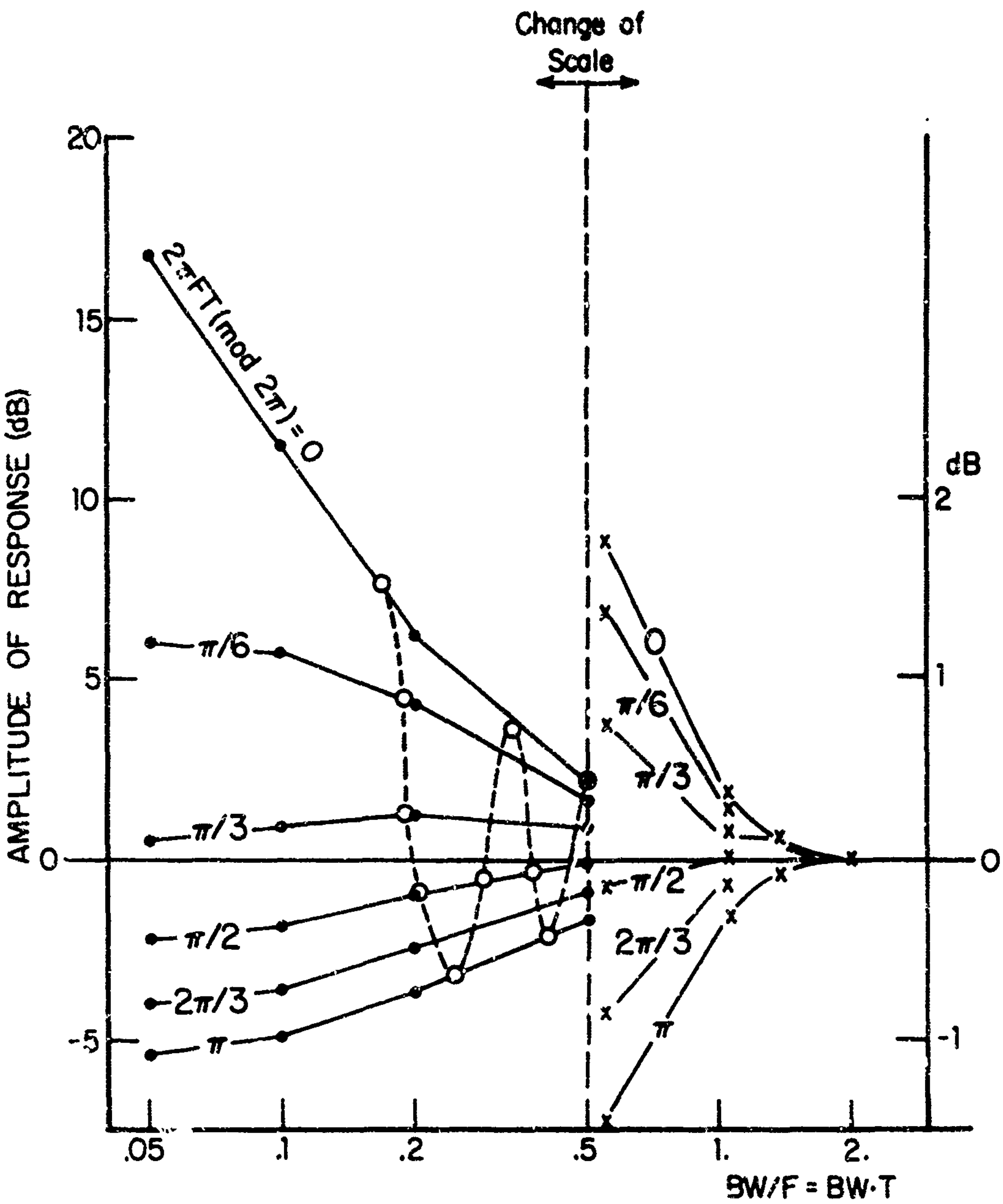


FIG. 4-3 HARMONIC OSCILLATOR RESPONSE

and 120 Hz. This curve is valid for all resonators having the same Q i.e., the same ratio of F to BW. However, the circles would represent different pulse rates. Thus this curve shows the behavior for F = 600 and Bw = 100, but all the pulse frequencies cited should be doubled.

The extent of the response change from maximum to minimum can be reduced if the driving pulses occur at intervals which are alternately shorter and longer. This is discussed in the next section.

4.2 Alternate Pulsing of A Resonator

In commenting on the appreciable change in the response of a resonator, we imply that perhaps the resulting maxima or minima are undesirable features of our model of speech production which actually do not exist because of some physical or neurological mechanism in the actual human speech production system. We thus are interested in simple models for reducing the height of the maxima or depth of the minima. As will be shown in what follows, the replacement of the constant period pulse source by one whose pulses occur at alternately short and long intervals gives such a reduction. The interest in such a model is increased as a result of the obscuration that such alternations actually occur in human speech (Lieberman, 1961; Smith, 1968). In the discussion that follows we will discuss alternation as a means of increasing the response during what would otherwise be minima. Such a discussion is based on a premise that optimal speech production is that with the greatest amplitude. The alternative is that alternation works to lower response maxima which also correspond to maxima of the impedance presented to the larynx by the vocal tract. While we do not orient our discussion to this latter case all the same principles apply and the same equations may be used to measure the effect.

A model for the generation of alternating pulses is shown in Fig. 4.4. A pulse generator, operating at a rate equal to half the number of pulses per second we desire, drives a linear system whose

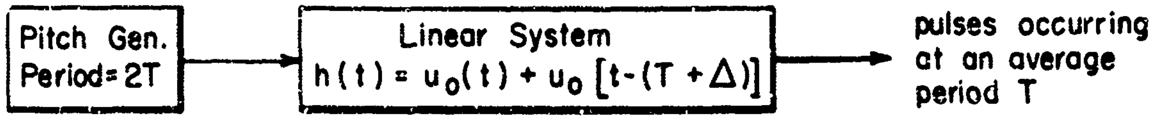


FIG. 4.4 MODEL FOR GENERATION OF ALTERNATING PERIOD PULSES

output is two pulses for every pulse in. The pulses occur at the average rate we desire and have inter-pulse intervals which alternate between $(T+\Delta)$ and $(T-\Delta)$ seconds. The affect of the alternation may be seen by considering how the frequency components of the pulse generator are affected by the filter.

The frequency components occur at multiples of $\frac{1}{2T}$ as is indicated by the vertical lines in Fig. 4.5. The transfer function of the dual pulse filter is

$$H(j\omega) = 1 + e^{-j\omega(T+\Delta)} \tag{4.7}$$

The magnitude of this transfer function is

$$|H(j\omega)| = \left|\cos\left[\frac{\omega}{2}(T + \Delta)\right]\right| * 2 \tag{4.8}$$

The effect of different values of Δ can be seen in Fig. (4.5) where the magnitude of the transfer function is plotted for $\Delta = 0$ and $\Delta = T/4$.

For $\Delta = 0$, the cosine function cancels all the odd components. The resulting even harmonics are actually all the harmonics of a pulse train of rate $\frac{1}{T}$. This is actually the case because, without the Δ, we do have a constant period pulse train with period T. For $\Delta = T/4$ we do, however, pass with maximum magnitude one of the odd components while suppressing its even neighbors. Thus if the peak of the resonance were at A in Fig. 4.5 there would be no need to alternate the pulses. This corresponds to the situation shown by the dashed lines in Fig. 4.1, depicting a component occurring at the resonance peak. The situation depicted by the solid lines in Fig. 4.1 corresponds to the resonance peak occurring at B in Fig. 4.5, half way between components of the average pulse frequency. In this case, a Δ of T/4 changes what would otherwise be a minimum response condition to a maximum by generating a maximum component at the resonance peak. For peaks which occur at frequencies which are not multiples of 1/2T, the maximum response can be obtained by finding the component which is nearest to the peak

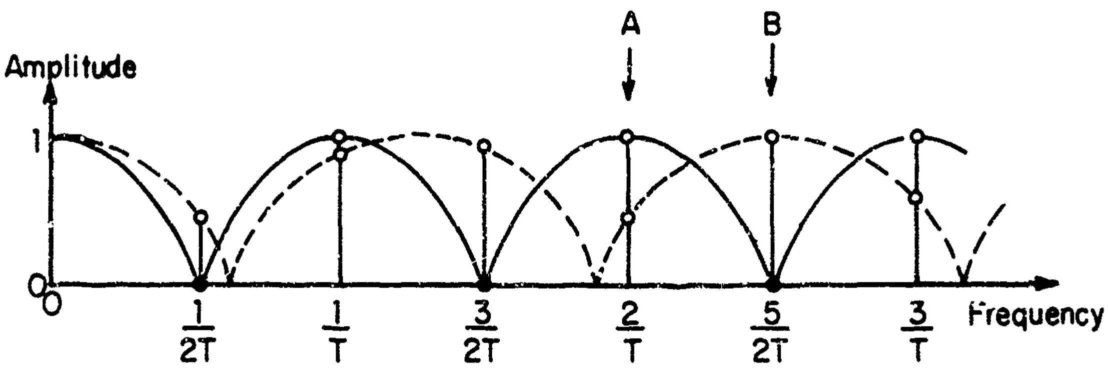


FIG. 4-5 EFFECT OF JITTER ON COMPONENT AMPLITUDES

and maximizing it by the proper selection of Δ. This component is denoted as the kth component in the following formula. The formula is

$$\frac{\Delta}{T} = \begin{cases} \frac{1}{k} & \text{for } k \text{ odd} \\ 0 & \text{for } k \text{ even} \end{cases} \tag{4.9}$$

where k = integer value of $\left[2FT + \frac{1}{2}\right]$

and F is the frequency of the peak of the resonance.
The value of Δ is approximately half the reciprocal of the resonance frequency.

Theoretically, one could operate a maximum component arbitrarily close to a resonance peak. This is done by lowering the rate of the pulse generator in Fig. 4.4 and increasing the number of pulses in the impulse response of the filter by the same factor, to keep the average pulse rate the same. To maximize the proper component one would have to determine the correct timing for every pulse in the filter, by solving sets of transcendental equations. The complexity mediates against the model being representative of a natural process.

The complex signal representation used above for calculating the response to truly periodic pulses can also be used for the alternation situation. Here, however, we have two amplitudes: A_1 for the amplitude during the long period and A_2 for during the short. The formulas are best expressed as part of a descriptive table. To simplify the notation, we have set the amplitude of the excitation pulses to unity.

$$\xi(t) = \begin{cases} A_1 & \text{just after first pulse}[t=0 \pmod{2T}] \\ A_1 e^{ST+S\Delta} & \text{after long period} \\ A_2 = A_1 e^{ST+S\Delta} + 1 & \text{just after 2nd pulse} \\ A_2 e^{ST-S\Delta} = A_1 e^{2ST} + e^{ST-S\Delta} & \text{end of short period} \\ A_1 = A_1 e^{2ST} + e^{ST-S\Delta} + 1 & \text{just after 1st pulse} \end{cases}$$

Signal Amplitude	Instant of Time	
A_1	just after first pulse[t=0(mod 2T)]	
$A_1 e^{ST+S\Delta}$	after long period	$[t = T+\Delta]$
$A_2 = A_1 e^{ST+S\Delta} + 1$	just after 2nd pulse	$[t = T+\Delta]$
$A_2 e^{ST-S\Delta} = A_1 e^{2ST} + e^{ST-S\Delta}$	end of short period	$[t = 2T = 0 \pmod{2T}]$
$A_1 = A_1 e^{2ST} + e^{ST-S\Delta} + 1$	just after 1st pulse	$[t = 2T = 0 \pmod{2T}]$

where $S = \sigma + j\omega$ (4.10)

From Eq. (4.10) we obtain the equation for A_1

$$A_1 = \frac{1 + e^{ST-S\Delta}}{1 - e^{2ST}} \tag{4.11}$$

and by analogy

$$A_2 = \frac{1 + e^{ST+S\Delta}}{1 - e^{2ST}} \tag{4.12}$$

we also note that for $\Delta = 0$

$$A_1 = A_2 = \frac{1 + e^{ST}}{1 - e^{2ST}} = \frac{1}{1 - e^{ST}} \tag{4.13}$$

which is Eq. (4.3) for constant period pulse excitation

These expressions can now be used to derive some measure of response based on the two different response amplitudes. This most appropriate measure is probably the power averaged over the short and long intervals. The results of this calculation cannot be represented in simple graphical form as for the constant period case and is sufficiently complicated as to best be done for specific values of resonance frequency and bandwidth.

Such a comparison of resonator response power for alternated and constant period excitation is shown for three different resonator frequencies in Figs. 4.6 through 4.8. The horizontal axes show the average pulse frequency 1/T. On the average, the resonator power increases at 6 dB/octave following the input power from the constant amplitude excitation pulses. The curves for no alternation ($\Delta = 0$) show the same type of results given by House (1959). In determining the case for alternated pitch, the amount of alternation, Δ, was set to half reciprocal of the resonance frequency, rather than the reciprocal of the pitch component nearest the resonance frequency, as detailed above. As can be seen, this selection of Δ makes the response to alternated pulses be 180° out of phase with the response to constant period pulses. One has peaks where the other has valleys and vice-versa. Thus for any combination of resonator and pitch frequencies, resonator response may be either maximized or minimized by selection of the proper pitch mode: alternated or constant period.

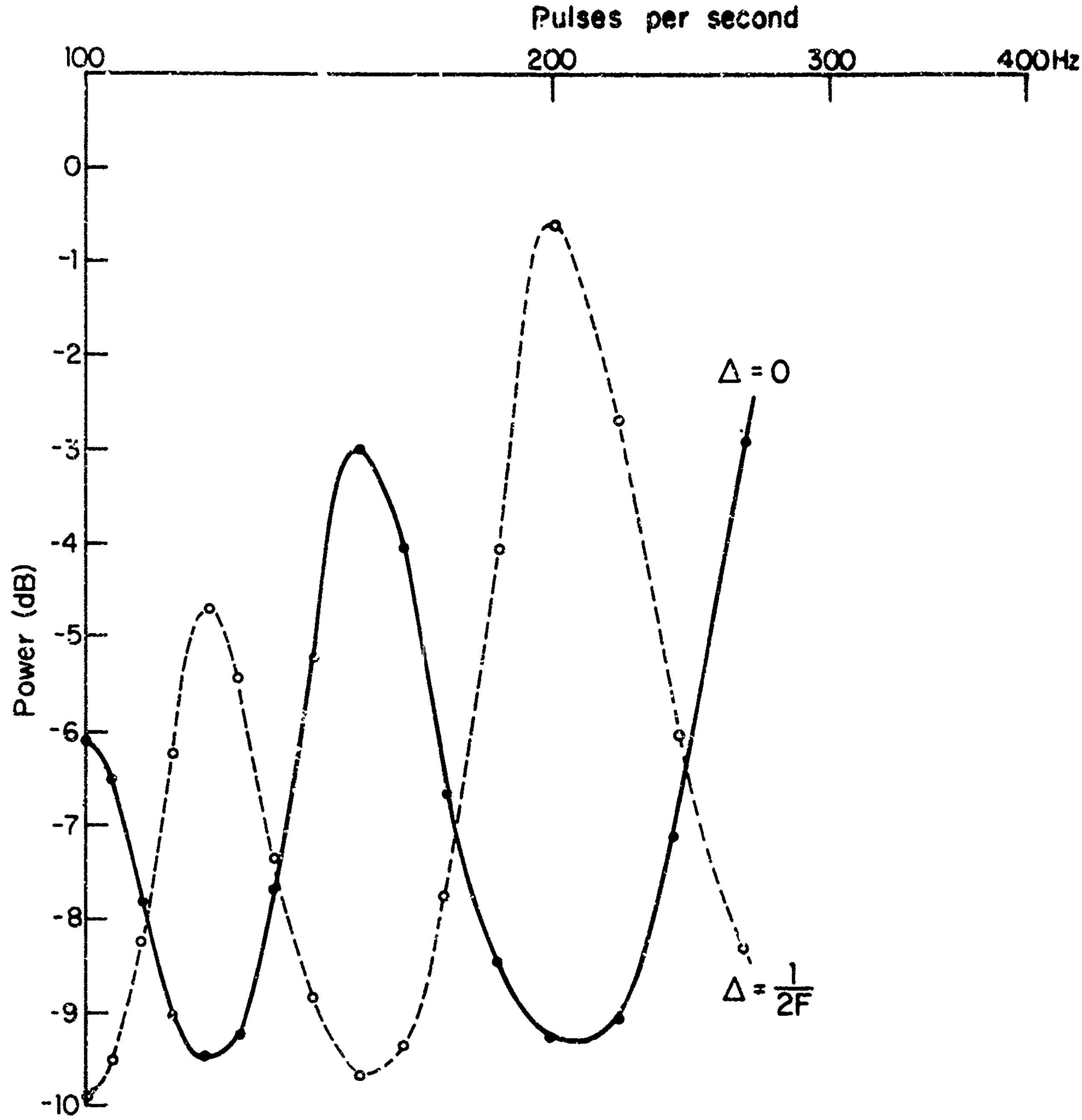


FIG. 4-6 RESPONSE POWER FOR ALTERNATED AND CONSTANT PERIOD PULSES EXCITING A RESONATOR OF F = 300 Hz, BW = 50 Hz

SIGNATRON®

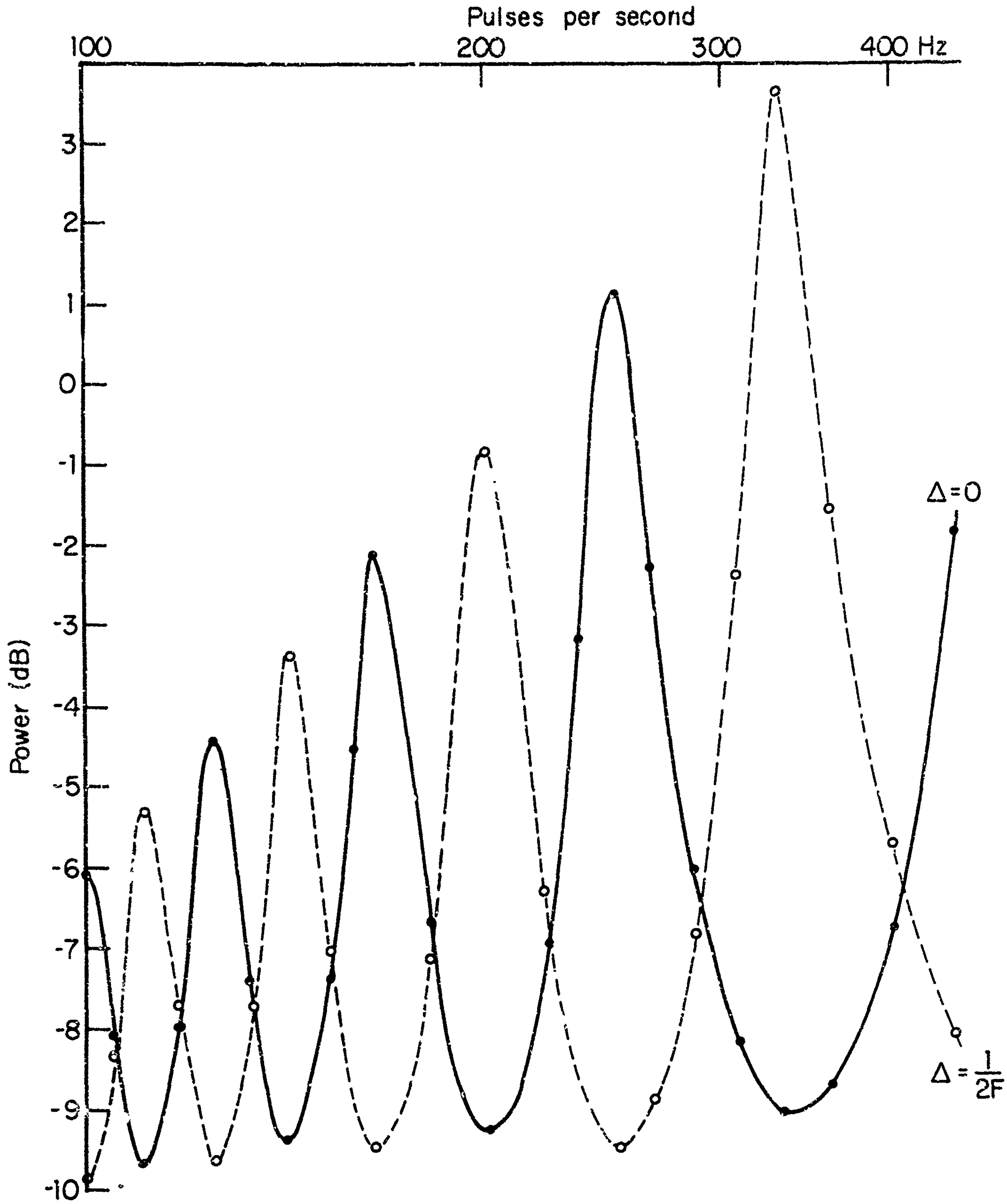


FIG. 4-7 RESPONSE POWER FOR ALTERNATED AND CONSTANT PERIOD PULSES EXCITING A RESONATOR OF F=500 Hz, BW=50 Hz.

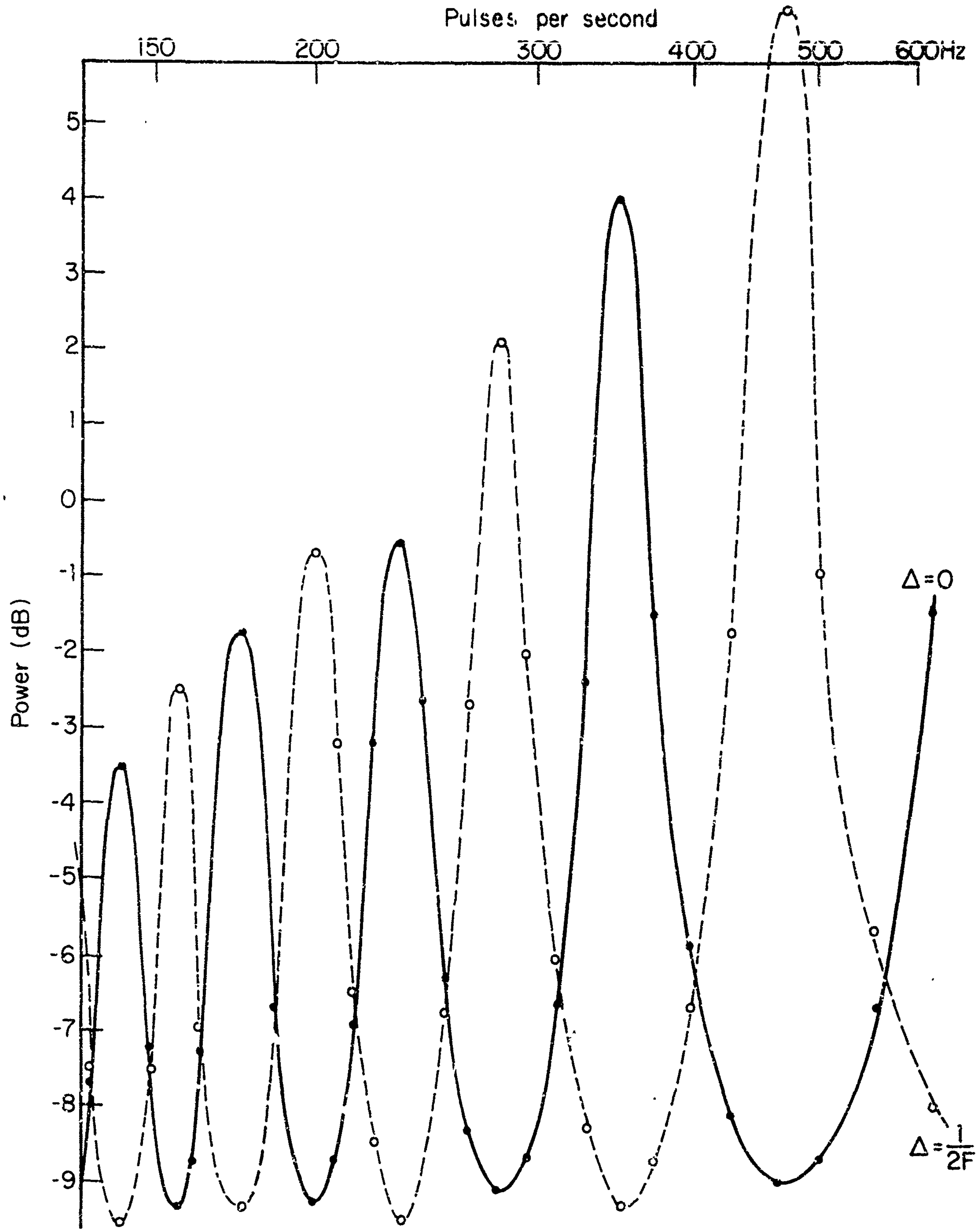


FIG. 4-8 RESPONSE POWER FOR ALTERNATED AND CONSTANT PERIOD PULSES EXCITING A RESONATOR OF F=700 Hz, BW=50 Hz

REFERENCES

Dunn, H. K.: Methods of Measuring Vowel Formant Bandwidths, J. Acoust. Soc. Am. 33, 1737-1746 (1961).

Fant, G.: Acoustic Theory of Speech Production, 's-Gravenhage: Mouton & Co. 1960.

Flanagan, J. L.: A Difference Limen for Vowel Formant Frequency. J. Acoust. Soc. Am. 27, 613-617 (1955).

Flanagan, J. L.: Speech Analysis, Synthesis and Perception. New York: Academic Press, Inc. 1965.

Holmes, J. N.: An Investigation of the Volume Velocity Waveform at the Larynx during Speech by Means of an Inverse Filter. Proc. IV Int. Congress Acoust., Copenhagen, Denmark, August 1962. Also Proc. Stockholm Speech Comm. Seminar, RIT, Stockholm, Sweden, September 1962.

House, Arthur S.: A Note on Optimal Vocal Frequency. J. Speech and Hearing Res., 2, 55-60 (1959).

Lieberman, P.: Perturbations in Vocal Pitch. J. Acoust. Soc. Am. 33, 597-603 (1961).

Lindqvist, J.: Inverse Filtering -- Instrumentation and Techniques. STL-QPSR-4/1964, Speech Transmission Lab., Royal Inst. of Tech., Stockholm. 1-4, (1964).

Lindqvist, J.: Studies of the Voice Source by Means of Inverse Filtering. STL-QPSR-2/1965, Speech Transmission Lab., Royal Institute of Tech., Stockholm. 8-13 (1965).

Mathews, M. V., J. E. Miller, and E. E. David, Jr.: An Accurate Estimate of the Glottal Waveshape. J. Acoust. Soc. Am. 33, 843(a) (1961).

Peterson, G. E., and H. L. Barney: Control Methods Used in a Study of the Vowels. J. Acoust. Soc. Am. 24, 175-184 (1952).

Smith, C. P.: Private Communication (1968).

Stevens, K. N.: The Perception of Sounds Shaped by Resonance Circuits. ScD Thesis, Massachusetts Institute of Technology. Cambridge, Mass., 1952.

Appendix A

INSTRUCTION MANUAL FOR VOTIF FILTERING UNITS

Prepared by:

Design Automation, Inc.
809 Massachusetts Avenue
Lexington, Massachusetts 02173

Prepared for:

SIGNATRON, Inc.
594 Marrett Road
Lexington, Massachusetts 02173

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

INSTRUCTION MANUAL FOR FILTERING INSTRUMENT

1.0 Introduction

This appendix describes the design and operation of the Null and Resonance Filters of the VOTIF speech analyser. Operational instructions are given for a composite filtering instrument which consists of five Null Filters and one Resonance Filter connected in cascade. The frequency and bandwidth of each of these filters may be set independently over the tuning range shown in Figure 1. Each filter operates independently of the other filters.

The instrument operates in 50°F to 125°F ambient temperature without forced-air cooling, and operates from a standard 117 VAC 60-Hz commercial power line. A two-section 19-inch rack-mounting frame contains the instrument input and output BNC connector clusters, a regulated dual-output power supply, and quick-disconnect $\frac{1}{4}$-turn panel-mount fasteners for mounting all six filter units in the frame. Shielded cables with BNC connectors are furnished for interconnection of filter units. The power supply is an Acopian Model 15D70U rated for dual 15V 700 mA operation.

Figure 2 shows an appropriate installation arrangement for the units. Various factors discussed in subsequent sections affect the actual arrangement used in any given analysis situation. In all situations it is advisable to have the lowest noise units earliest in the chain to minimize noise build-up. This noise build-up is a consequence of the rising gain-frequency characteristic (12 dB/octave/null) of the instrument. For the maintenance of highest output signal-to-noise ratio, the null units should be adjusted so that the tuning frequencies increase along the cascade with the first unit having the lowest frequency setting. However, when the input signal is noisy, as is often the case with speech signals, the reverse ordering may be more advisable. While not keeping signal-to-noise ratio to a minimum, having tuning frequencies decrease along the cascade will tend to minimize noise levels at each stage of the cascade.

Because any imperfections of the signal source will be magnified by the rising gain-frequency response characteristic of the instrument, it is suggested that precautions be taken to minimize distortion, pickup and noise in the input signal. Similarly, when the output of a sine-wave signal generator is used as a test input signal, imperfections in the signal generator output that are barely visible on an oscilloscope trace will be magnified by the rising gain-frequency response of a Null Filter. Many sine-wave signal generators (including the Hewlett-Packard Model 209A) have small discontinuities at the sine-wave zero-crossings. These will be accentuated in the Null Filter, resulting in narrow spikes at the sine-wave zero-crossings. This effect is most easily seen at TP5 in the Null Filter. Another imperfection of some signal generators is the presence of random noise added to the signal <u>after</u> the output level control. When the generator output is set to minimum, the output noise will still remain. Thus, when testing the internally-generated noise of the instrument, the instrument input should be physically shorted to remove noise which could be coming from the signal source.

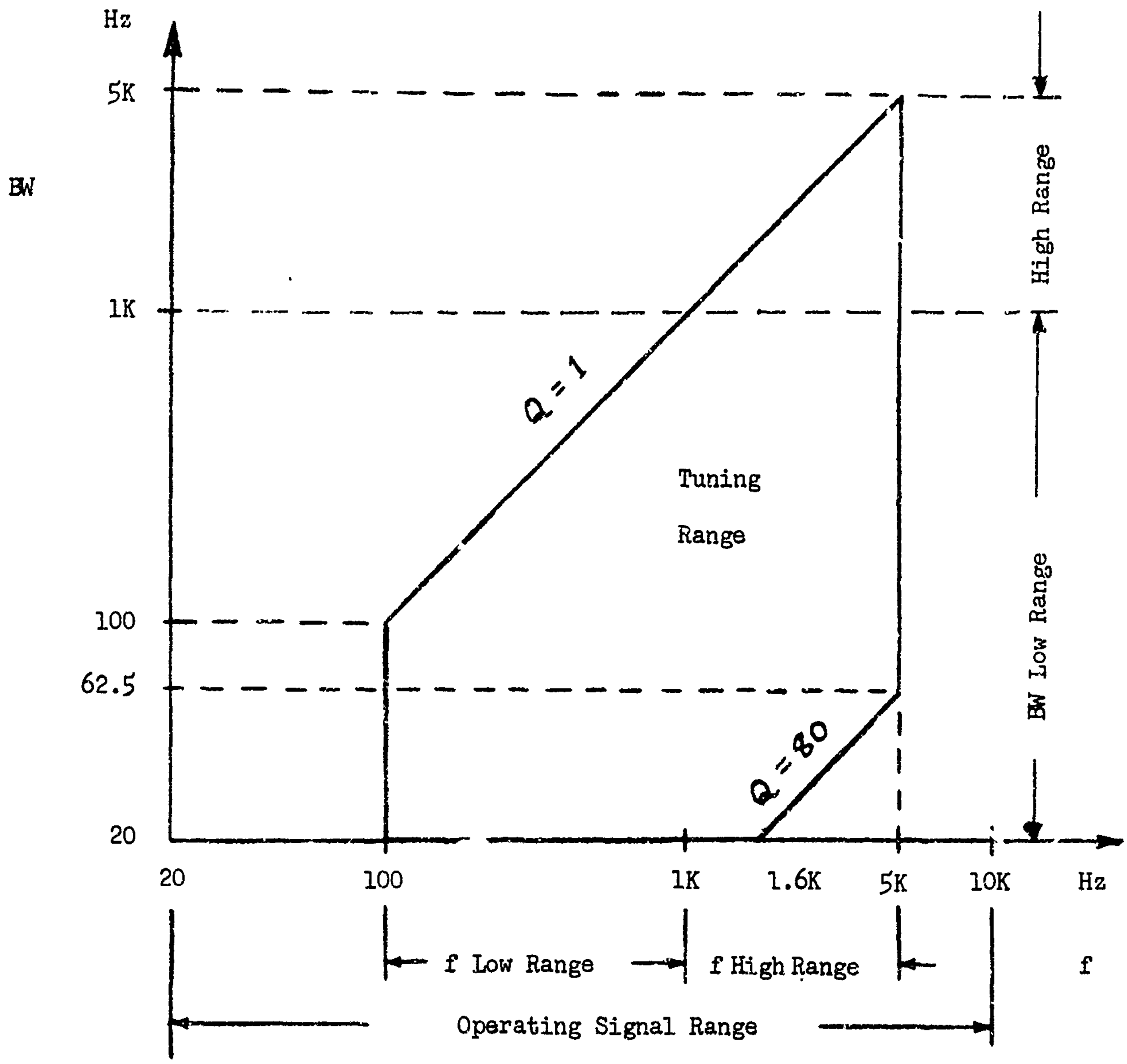


Figure 1. Tuning Range of Frequency and Bandwidth Control Settings

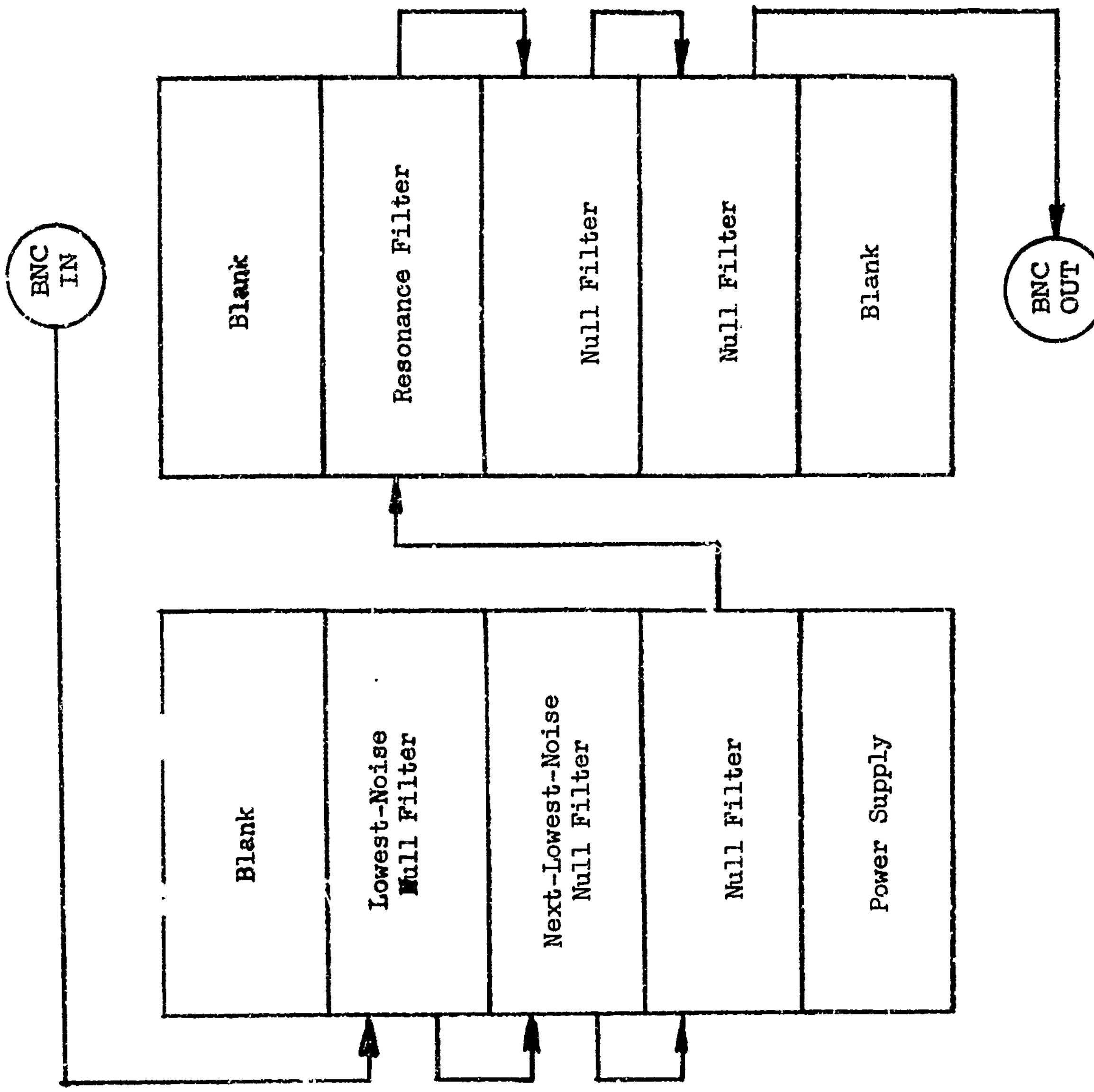


Figure 2. Recommended Installation Arrangement

2.0 Null Filter Functional Description

The Null Filter has a target transfer function which represents a second-order anti-resonance or Null Filter with unity effective DC gain, and is given by

$$H_1\ (S) = \frac{(s + b)^2 + a^2}{b^2 + a^2}$$

The filter frequency and bandwidth parameters, a and b respectively. are independently tunable over the audio frequency range by means of precision dials calibrated in Hertz (cps).

Modifications to the above transfer function incorporated into the design comprise an 18 KHz low-pass filter for roll-off of overall high-frequency response, roll-off of the s^2 term at 100 KHz. and polarity inversion (negative sign) of the effective DC gain (extrapolation of the low-frequency gain to DC).

2.1 Null Filter Specification Summary

2.1.1. Controls

IN-OUT Switch	IN: Output BNC connected to Input BNC OUT: Output BNC connected to filter output
GAIN Control	Adjusts overall gain through filter, after setting FREQ
BW Control and Range Switch	LOW range: 100 Hz/turn, up to 1000 Hz HIGH range: 1 KHz/turn, up to 5 KHz Limits: As defined in Fig. 1
FREQ Control and Range Switch	LOW range: 100 Hz/turn, up to 1000 Hz HIGH range: 1 KHz/turn, up to 5 KHz Limits: As defined in Fig. 1

2.1.2 Accuracy

FREQ Dial	Adjustment precision: ±0.5% of value Calibration accuracy: ± 2% of value
BW Dial	Adjustment precision: ± 0.5% of FREQ for FREQ ≥ 100 Hz min., otherwise ± 0.5 Hz Calibration accuracy: ± 10% of value
Transfer Function	Signal operating range: 20 Hz to 10 KHz Relative amplitude: ± 0.25 dB (± 2.9%) Delay variation: ± 0.10 msec

2.1.3 Impedance Levels

Input	2.2 kilohms ± 5%, capacitor-coupled
Output	2 ohms typical
Rated Load	2 kilohms minimum impedance

2.1.4 Signal Levels

Output	Up to ± 10V peak into 2 kilohms minimum load impedance, for sine-wave signals of ≥ 200 Hz on LOW FREQ and ≥2 KHz on HIGH FREQ. Below these frequencies, maximum output is determined by internal signal level at TP5 or TP6, and is a function of FREQ and BW control settings.
Input	Up to value causing maximum output; varies with GAIN, FREQ and BW settings and input frequency. The proper input signal level and GAIN setting are discussed in Section 2.2.

2.1.5. Noise Level — At least 40 dB below 7 Vrms at output; improves with increasing FREQ setting.

2.1.6 Test Points

All test points are isolated by resistors of 680 or 1000 ohms to prevent damage in case of accidental shorting of a test point to ground. The test points are:

TP1	Input connector
TP2	Spare
TP3	Differentiator channel output
TP4	Bandwidth channel output
TP5	Summing amplifier output (unfiltered)
TP6	Input amplifier output
TP7	+ 15V supply
TP8	- 15V supply
TP9	Frequency channel output
TP10	Output connector

2.1.7 Power Drain

No-signal	69 mA at + 15V, -72 mA at -15V
Normal signals	89 mA at + 15V, -92 mA at -15V

2.2 Null Filter Operating Instructions

An appropriate installation arrangement for the Null Filter is shown in Figure 2. Each filter mounts and dismounts by means of $\frac{1}{4}$-turn panel fasteners, and is connected by means of BNC signal input and output connectors and a multi-pin power connector in the rear.

Front-panel control functions, dial calibrations and operating limits, and test-point functions are listed in the Specification Summary. After the FREQ and BW dials have been set, the GAIN may be set as high as the value that gives unity effective DC gain. This value is obtained when the output amplitude of low-frequency signals (20 Hz) is unaffected by IN-OUT Switch operation.

If the GAIN setting or input signal level is too high, saturation or other distortion may occur. If the input signal level is too low, signal-to-noise ratio may be reduced. Distortion conditions are best monitored at TP5, which precedes a low-pass filter followed by an output amplifier having a gain of ten. Signal and noise amplitudes are best monitored at TP10 which is connected to the output.

Choice of control settings should take account of signal-to-noise ratio, because in a cascade of Null Filter units the steeply rising gain-frequency characteristic (12 dB/octave per Null Filter) introduces significant noise gain and bandwidth. This rise reaches a peak at 18 KHz, where the low-pass filter in each Null Filter begins to roll off. In particular, it is recommended that the Null Filter GAIN controls be set at substantially less than unity effective DC gain (value discussed below). This will help to keep the high frequency noise of the first unit still moderately small at the output of the last unit. The noise gain and signal gain depend on FREQ and BW settings in all of the filter units.

To find a more desirable GAIN setting, let us assume that the 18 KHz noise content of the output of the first Null Filter is 5 mVrms. This passes through four Null Filters, one of which is approximately balanced out at 18 KHz by the Resonance Filter. Let us also assume that the final 18 KHz noise output should not exceed 1 Vrms. Then the 18 KHz gain of each Null Filter should be $\sqrt[3]{1/0.005} = 5.8$. This corresponds to unity gain at $18\sqrt{0.75/5.8} = 6.5$ KHz. The effective DC gain will be approximately $(\text{FREQ}/6.5\ \text{KHz})^2$, which is below unity by an amount dependent upon the FREQ setting. Thus the GAIN control can simply be set to obtain unity gain through each Null Filter at 6.5 KHz input signal frequency.

2.3 Null Filter Circuit Design

Figure 3 is a simplified transfer-function diagram of the Null Filter. For non-inverting input signals, the gain of an operational amplifier is larger by unity than the gain for inverting inputs. This fact is accounted for in Stage 4A, where both inputs are used, by means of the attenuation factor shown at the inverting input.

The simplified overall transfer function resulting from Figure 3 is as follows:

$$H(s) = -G_1 K_5 (T^2 s^2 + Ts(K_{4A} - 1)yK_{4B} + y^2 K_{4A} K_{4B} + x^2 K_3{}^2)$$

$$= -G_1 K_5 T^2 \left(s^2 + \frac{(K_{4A} - 1)K_{4B}\, ys}{T} + \frac{K_{4A} K_{4B}\, y^2}{T^2} + \frac{K_3{}^2 x^2}{T^2}\right)$$

We wish to realize the ideal transfer function:

$$H_1(s) = (s^2 + 2bs + b^2 + a^2) / (b^2 + a^2)$$

Let us define x and y as potentiometer transmissions (maximum = unity), f and BW as the dial readings in Hz, and F as the full-scale dial calibration of 10 kHz for both dials. We then have these relationships to be satisfied:

$$a = 2\pi f = 2\pi xF$$

$$b = \pi BW = \pi yF$$

$$2b = (K_{4A} - 1)K_{4B} y/T = 2\pi yF$$

$$b^2 = K_{4A} K_{4B} y^2/T^2 = \pi^2 y^2 F^2$$

$$a^2 = K_3{}^2 x^2/T^2 = (2\pi xF)^2$$

The last three equalities yield the design constraints:

$$(K_{4A} - 1)K_{4B} = 2\pi FT$$

$$(K_{4A} - 1)/K_{4A} = 2/\pi FT = 1 - 1/K_{4A}$$

$$K_3 = 2\pi FT$$

In this design the unity-gain frequency of the differentiator stages has been set to 2 kHz. This leads to the following design values:

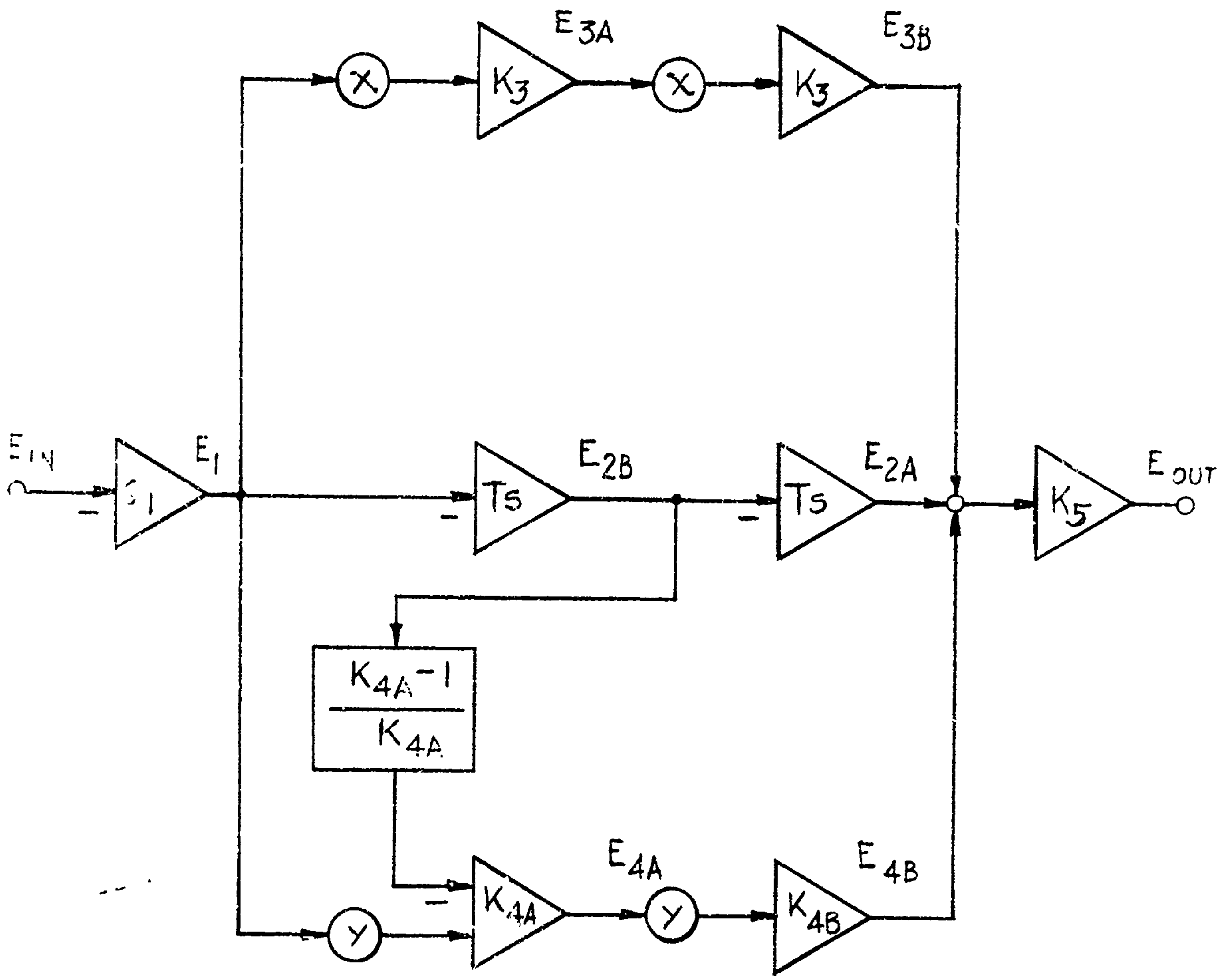


Figure 3. Simplified Transfer-Function Diagram of Null Filter

$T = 1/2\pi(2 \text{ kHz}) = 79.7$ usec

$K_3 = 5.0$

$K_{4A} = 5.0$

$K_{4B} = 1.25$

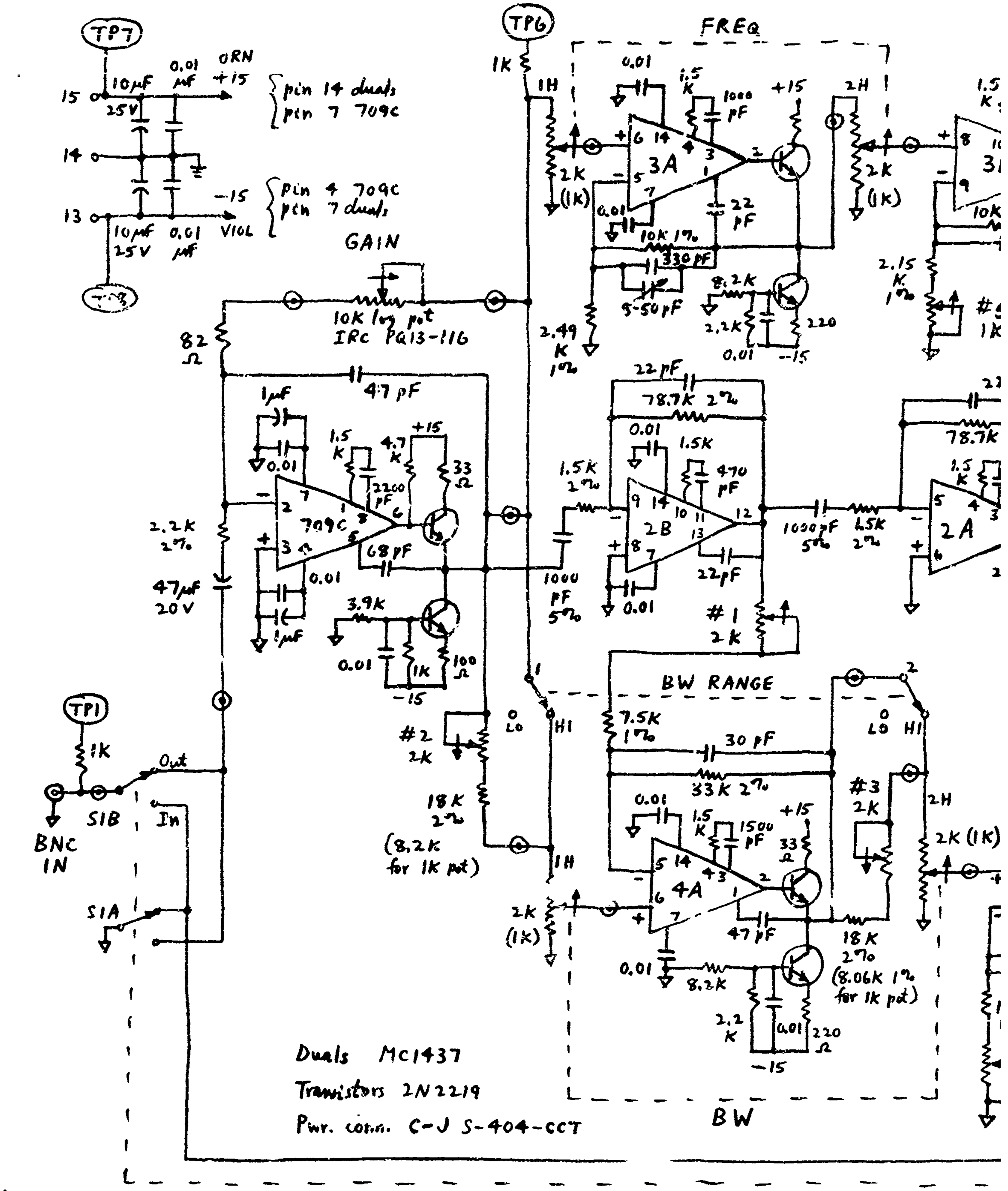
The variable gain control G_1 permits the gain factor $-G_1K_5T^2$ to be adjusted to meet the design requirement of unity effective DC gain. The gain G_1 would normally be varied inversely with $(a^2 + b^2)$. This factor can reach 5,000 : 1, which would use up much of the dynamic range available between noise and saturation levels if straight-forward range switching were used. This potential difficulty is largely avoided by the indirect method used for range switching. Ten-to-one range switching for both variables a and b is accomplished by scaling all other factors in the opposite direction. This is shown in the Circuit Schematic (Fig. 4). The effective DC gain is made insensitive to the Frequency Range Switch position. When the Frequency dial is maintained at one turn minimum by means of frequency range switching, the variation in G_1 is reduced to only 200 : 1. This permits reasonable signal-to-noise performance and together with a logarithmic infinite-resolution potentiometer aids manual gain adjustment.

Design factors which modify the transfer function above 10 kHz are the introduction of high-frequency rolloff in the differentiators and in the overall gain function. These rolloffs contribute to differentiator stability and to overall signal-to-noise ratio.

The s^2 term in the ideal transfer function corresponds to a gain-frequency asymptote rising at 12 dB/octave at the upper end of the operating signal frequency range (10 kHz). Above this point the frequency response must be rolled back to a falling asymptote for reasons of physical realizability, noise bandwidth limitation, and to maintain stability even in the presence of stray coupling.

Each differentiator stage has a pair of real poles at 100 kHz, producing only -0.1 dB and -12° at 10 kHz. The primary rolloff for the entire filter transfer function is provided by a fourth-order Butterworth low-pass filter at the output. With an 18-kHz cutoff frequency, the filter introduces only -0.1 dB with -87° at 10 kHz. Above its cutoff frequency, the fourth-order filter overrides the double differentiator, producing a net rolloff of 12 dB/octave up to 100 kHz. Beyond 100 kHz, each differentiator becomes -6 db/octave instead of + 6 dB/octave. The net rolloff beyond 100 kHz thus becomes 36 dB/octave.

Maximum overall gain occurs at the filter cutoff frequency, but does not exceed 24,000 over the entire range of dial settings. A net low-frequency gain inversion is utilized to make overall stability more insensitive to coupling from output to input. Stray coupling is minimized by physical separation and shielding of input and output leads, and by multiple bypassing and divided routing of power-supply lines.



SMALL POTS: 2K
LGE " 1K

POT SECT 1: FRONT (H=CW)
" " 2: REAR

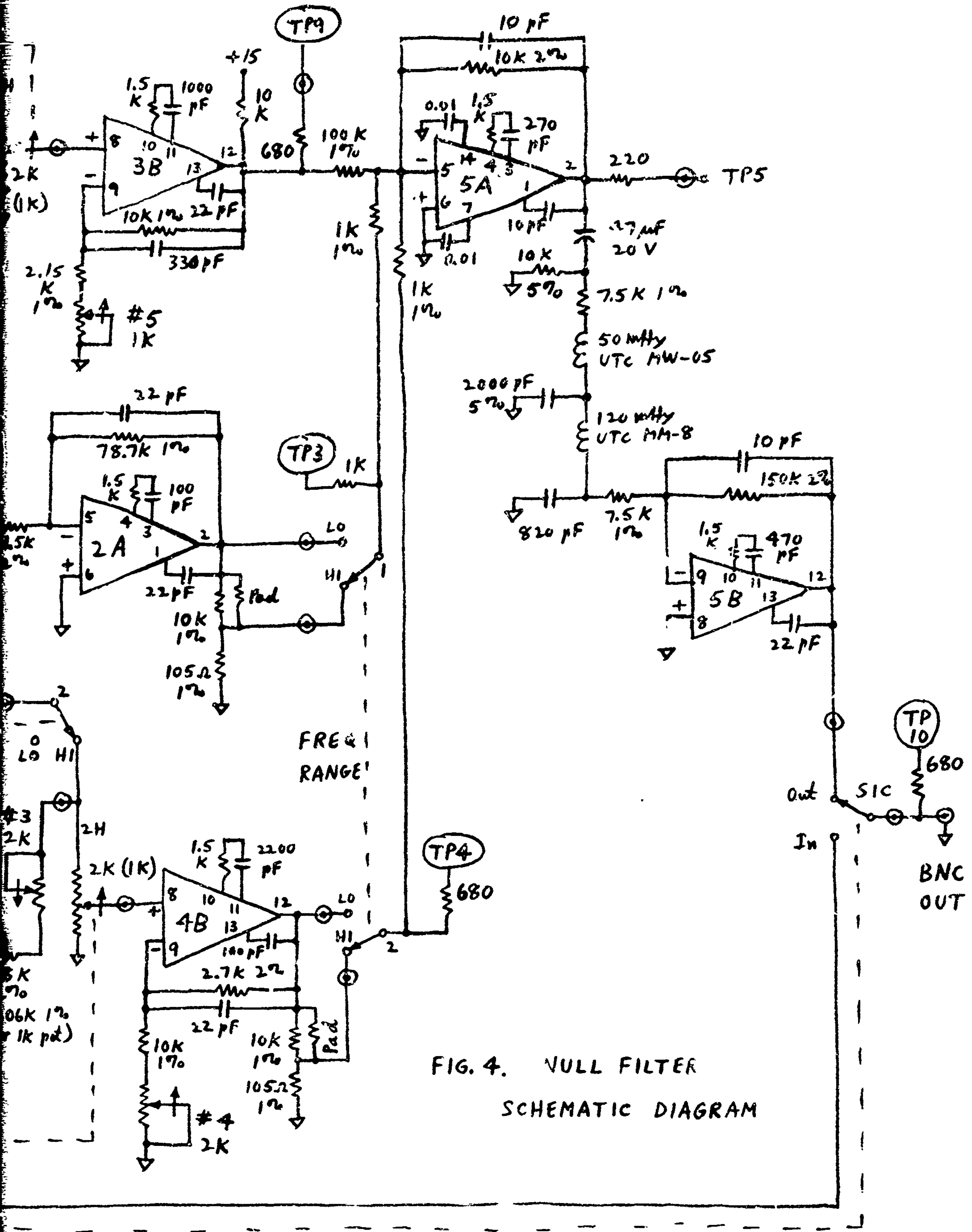


FIG. 4. NULL FILTER

SCHEMATIC DIAGRAM

RANGE SW. SECT 1 : TOP
" " " 2 : BOT

B

Each of the amplifier stages has a compensation network and rolloff feedback capacitor selected for accurate response to signal frequencies and effective discrimination against higher (noise) frequencies.

Emitter followers returned to current sources are used at two interstage locations for driving heavy loads with minimum amplifier crossover distortion.

The input amplifier is selected for low noise and is operated at low impedance levels to minimize the voltage output caused by the input current noise.

2.4 Null Filter Measured Response

The results of response measurements taken on Null Filter #1 on Nov. 1, 1968 are shown in Table 1. The frequency and bandwidth settings were 1000 Hz and 20 Hz, respectively, both on their low ranges. Measurements of both input and output voltage were made using a stable wideband full-wave operational rectifier feeding a Digitec DC digital voltmeter via a low-pass filter. Signal frequency of the Hewlett-Packard Model 209A oscillator was monitored with a Hewlett-Packard 512 frequency counter. The effective DC gain of the Null Filter was set close to unity, and the input or output, whichever was larger at each signal frequency, was set just below 7 V rms.

The measured null frequency was 1007 Hz, or 0.7% high, well within the ± 2% frequency calibration requirement. The ideal response data for use in Table 1 was computed for f = 1006 Hz and BW = 19 Hz for comparison with the actual frequency response.

The measured values were corrected for rectifier offset due to zero error and noise, and for rectifier amplitude non-linearity using a calibration curve. The ideal response was normalized to the measured low-frequency gain to eliminate the effect of the slight difference from unity in the effective DC gain.

Table 1. Measured Response at 1000 Hz Frequency and 20 Hz Bandwidth Dial Settings.

Signal Frequency f_s Hz	Measured Response H_M	Ideal Response H_I (f=1006 Hz, BW=19 Hz)	Ratio H_M/H_I	Error dB
40	0.99842	0.99842	1.000	0
700	0.52151	0.51604	1.011	0.10
741	0.46260	0.45771	1.011	0.10
823	0.33587	0.33114	1.014	0.12
864	0.26570	0.26295	1.010	0.09
935.7	0.14087	0.13609	1.035	0.30
966.5	0.08297	0.07917	1.048	0.41
993	0.03280	0.03180	1.031	0.27
1007	0.01704	0.01900	0.897	-0.94
1019	0.03280	0.03222	1.018	0.16
1058	0.11050	0.10780	1.025	0.21
1200	0.42799	0.42335	1.011	0.10
1452	1.0955	1.0834	1.011	0.10
1757	2.0635	2.0503	1.006	0.05
2572	5.5408	5.5362	1.001	0.01
4143	15.598	15.959	0.977	-0.20

The results in Table 1 show that the relative amplitude limit of ±0.25 dB is met at all the test frequencies except at and near the null, where the response is down 20 to 35 dB. The largest error occurs right at the null. It is believed that these errors are caused primarily by measuring instrument non-linearity and zero offset, which are large enough to require a more accurate linearity calibration of the rectifier, together [illegible] a rectifier range switching arrangement, to resolve definitely the [illegible] for the apparent disagreement between measured and ideal responses near the deep null. The response was deemed to be close enough to the ideal not to warrant development of more precise instrumentation.

Bandwidth dial calibration was checked by taking measurements with settings of 1000 Hz frequency and 200 Hz bandwidth ($Q = 5$). With unity nominal effective DC gain, the measured response was 0.1988 at 1000 Hz and 0.9927 at 100 Hz, giving a ratio of 0.2003. This is within 0.1% of the ideal ratio 0.1983/0.9903 = 0.2002, or two orders of magnitude better than the ± 10% bandwidth calibration specification.

Noise output measurements taken on the same date are shown in Table 2. The effective DC gain was set to unity for each tuning frequency, and the bandwidth was set at zero. The input was shorted, representing low impedance of the input signal source. The output readings were corrected for rectifier zero offset and converted to rms values. The noise output is highest at the lowest tuning frequency, where the transfer function response up to and including 18 kHz is largest. Using the maximum available output signal of 7 V rms as a reference, the signal-to-noise ratio is 51 dB or better, substantially better than the required 40 dB.

Table 2. Measured Noise Output with Effective DC Gain Set to Unity at Various Tuning Frequencies.

Frequency Setting f Hz	Frequency Range	Noise Output vrms	Level Referred to 7Vrms dB
100	low	0.020	-51
200	low	0.0055	-62
1000	high	0.0022	-70
2000	high	0.006	-61

2.5 Null Filter Maintenance and Calibration

Stability of performance of the Null Filter is safeguarded by means of adequate design margins and frequency compensation techniques, careful component and wiring layout and shielding, and the use of stable metal-film resistors and trim potentiometers. Critical capacitors are stable low-loss mica types, and the input amplifier is a selected low-noise 709C.

Should it be necessary to replace any components, consideration should be given, after the repair is completed, as to whether the gain of a critical stage (and therefore the overall calibration) might be affected. This applies primarily to resistors connected to the input terminals of amplifiers preceding the three-input summing amplifier. Examination of the Factory Calibration Procedure below should enable determining which, is any, calibration steps are affected.

Recalibration due to aging or drift should not be necessary for at least a year. A simple way to verify stability is to check null frequency at several points at near-zero bandwidth, using a signal generator and a frequency counter.

Below is the Factory Calibration Procedure, which utilizes a DC digital voltmeter to set gain and attenuation ratios within 0.2% accuracy. Refer to the schematic of Figure 4.

Factory Calibration Procedure

1. Check alignment of electrical zero of each section of FREQ and BW pots to dial zero, using an ohmmeter.

2. Set trimmer #1 to obtain gain = -4 from 2B output to 4A output. Set BW = 0, and obtain 2 VDC at 2B output by means of GAIN pot and jumpers connecting 47 uF negative end to -15V and 47 k ohm across 1000 pF feeding 2B.

3. Set trimmer #2 to obtain 10:1 ratio at BW pot 1H terminal with BW Range switching. Use BW = 0, and 10 VDC at output of 709C stage.

4. Set trimmer #3 to obtain 10:1 ratio at BW pot 2H terminal with BW Range switching. Use BW = 0, and 10 VDC at output of stage 4A.

5. Set trimmer #4 to obtain gain = 5/4 through stage 4B. Use BW = approximately 7000 (high range) and adjust GAIN to obtain 8 VDC at + input of stage 4B. Set FREQ = 0 (high range).

6. Pad 10K 1% resistor at output of stage 4B to obtain 100:1 ratio at arm 2 of FREQ RANGE switch between high and low positions. Use 10 VDC at 4B output, and check that grounding - input of stage 5A has no effect.

7. Pad 10K 1% resistor at output of stage 2A to obtain 100:1 ratio at arm 1 of FREQ RANGE switch, as above.

8. Check for unity gain through stages 2B and 2A at 2000 Hz input frequency, and trim 78.7K resistor or 1000 pF capacitor if necessary.

9. Set trimmer #5 for best null at FREQ = 1000 Hz (low range), BW = 0 (low range), and with 1000 Hz input signal. Check FREQ scale reading for best null at 500 Hz input.

10. Adjust variable capacitor at stage 3A for best null at FREQ = 5 kHz (high range), BW = 0 (low range) and 5 kHz input. Check FREQ scale reading for 2 kHz and 1 kHz input signals.

3.0 Resonance Filter Functional Description

The Resonance Filter has a target transfer function which is the inverse of the Null Filter target transfer function. It is given by

$$H_1(s) = \frac{b^2 + a^2}{(s + b)^2 + a^2}$$

The Filter frequency and bandwidth parameters, a and b respectively, are independently tunable over the audio frequency range by means of precision dials calibrated in Hertz (cps).

The only modification to the above transfer function included in the design is the inverted polarity (negative sign) of the effective DC gain.

3.1 Resonance Filter Specification Summary

3.1.1 Controls

IN-OUT Switch	IN: Output BNC connected to Input BNC OUT: Output BNC connected to filter output
GAIN Control	Adjusts overall gain through filter, after setting FREQ
BW Control and Range Switch	LOW range: 100 Hz/turn, up to 1000 Hz HIGH range: 1 KHz/turn, up to 5 KHz Limits: As defined in Fig. 1
FREQ Control and Range Switch	LOW range: 100 Hz/turn, up to 1000 Hz HIGH range: 1 KHz/turn, up to 5 KHz Limits: As defined in Fig. 1

3.1.2 Accuracy

FREQ Dial	Adjustment precision: ±0.5% of value Calibration accuracy: ± 2% of value
BW Dial	Adjustment precision: ± 0.5% of FREQ for FREQ ≥ 100 Hz min., otherwise ± 0.5 Hz Calibration accuracy: ± 10% of value
Transfer Function	Signal operating range: 20 Hz to 10 KHz Relative amplitude: ± 0.25 dB (± 2.9%) Delay variation: ± 0.10 msec

3.1.3 Impedance Levels

Input	3 to 10 kilohms, capacitor-coupled
Output	2 ohms typical
Rated Load	2 kilohms minimum impedance

3.1.4 Signal Levels

Output	Up to ± 10V peak into 2 kilohms minimum load impedance. At some control settings, maximum output is determined by internal signal levels, by the requirement of keeping internal levels at or below ± 10V.
Input	Up to value causing distortion at TP5; varies with GAIN setting.

3.1.5 Noise Level

At least 40 dB below 7Vrms at output

3.1.6 Test Points

All test points are isolated by resistors of 680 or 1000 ohms to prevent damage in case of accidental shorting of a test point to ground. The test points are:

TP1	Input connector
TP2	Spare
TP3	Spare
TP4	Spare
TP5	Frequency feedback channel output
TP6	Spare
TP7	+ 15V Supply
TP8	- 15V Supply
TP9	Spare
TP10	Output connector

3.1.7 Power Drain

No-signal	22 mA at + 15V, -22 mA at -15V
Normal signals	41 mA at + 15V, -41 mA at -15V

3.2 Resonance Filter Operating Instructions

An appropriate mounting location for the Resonance Filter is shown in Figure 2. Filter mounting and connection are the same as described for the Null Filter.

Front-panel control functions, dial calibrations and operating limits are the same as for the Null Filter. They are listed in the Specification Summary together with other parameters of the Resonance Filter.

It is recommended that the GAIN control be set for unity effective DC gain. The considerations which affect the choise of Null Filter GAIN control setting, discussed in Section 2.3, need not be considered here because the gain of the Resonance Filter falls at frequencies beyond resonance rather than rising like the Null Filter. Thus noise is attenuated, rather than amplified, and the GAIN control can be set for unity effective DC gain.

Test point TP5 is provided to aid in detecting saturation or distortion conditions due to excessive input signal level. Both TP5, the output of a limiting amplifier, and TP10, the Filter output, should be monitored for this purpose. Proper signal-to-noise ratio resulting from adequate input signal level would be observed at TP10.

3.3 Resonance Filter Circuit Design

Figure 5 shows a simplified transfer-function diagram of the Resonance Filter, which uses feedback combined with feed-forward through two integrating amplifiers.

The net input to the summing amplifier can be expressed as follows:

$$e_s = G_1E_{in} + K_By\,(K_1y + C_IT_2s)\,E_{I2} + C_FK_FK_1x^2E_{I2}$$

$$e_s = -T_1T_2s^2E_{I2}/K_s$$

Eliminating e_s, we can obtain a ratio of the variables:

$$-\frac{G_1E_{in}}{E_{I2}} = \frac{T_1T_2s^2}{K_s} + K_ByC_IT_2s + K_1K_By^2 + C_FK_1K_Fx^2$$

This expression is then used in the overall transfer function:

$$H\,(s) = \frac{E_{out}}{E_{in}} = \frac{K_oK_1xE_{I2}}{E_{in}}$$

$$H\,(s) = \frac{-G_1K_oK_1xK_s/T_1T_2}{s^2 + \frac{K_sK_BC_Iys}{T_1} + \frac{K_sK_1K_By^2}{T_1T_2} + \frac{C_FK_sK_1K_Fx^2}{T_1T_2}}$$

This will represent the ideal transfer function:

$$H_1\,(s) = \frac{b^2 + a^2}{s^2 + 2bs + b^2 + a^2}$$

Using the same definitions as Section 2.3, we obtain the following relationships:

$$a = 2\pi f = 2\pi xF$$

$$b = \pi BW = \pi yF$$

$$2b = K_sK_EC_Iy/T_1 = 2\pi yF$$

$$b^2 = K_sK_1K_By^2/T_1T_2 = \pi^2y^2F^2$$

$$a^2 = K_sC_FK_1K_Fx^2/T_1T_2 = 4\pi^2x^2F^2$$

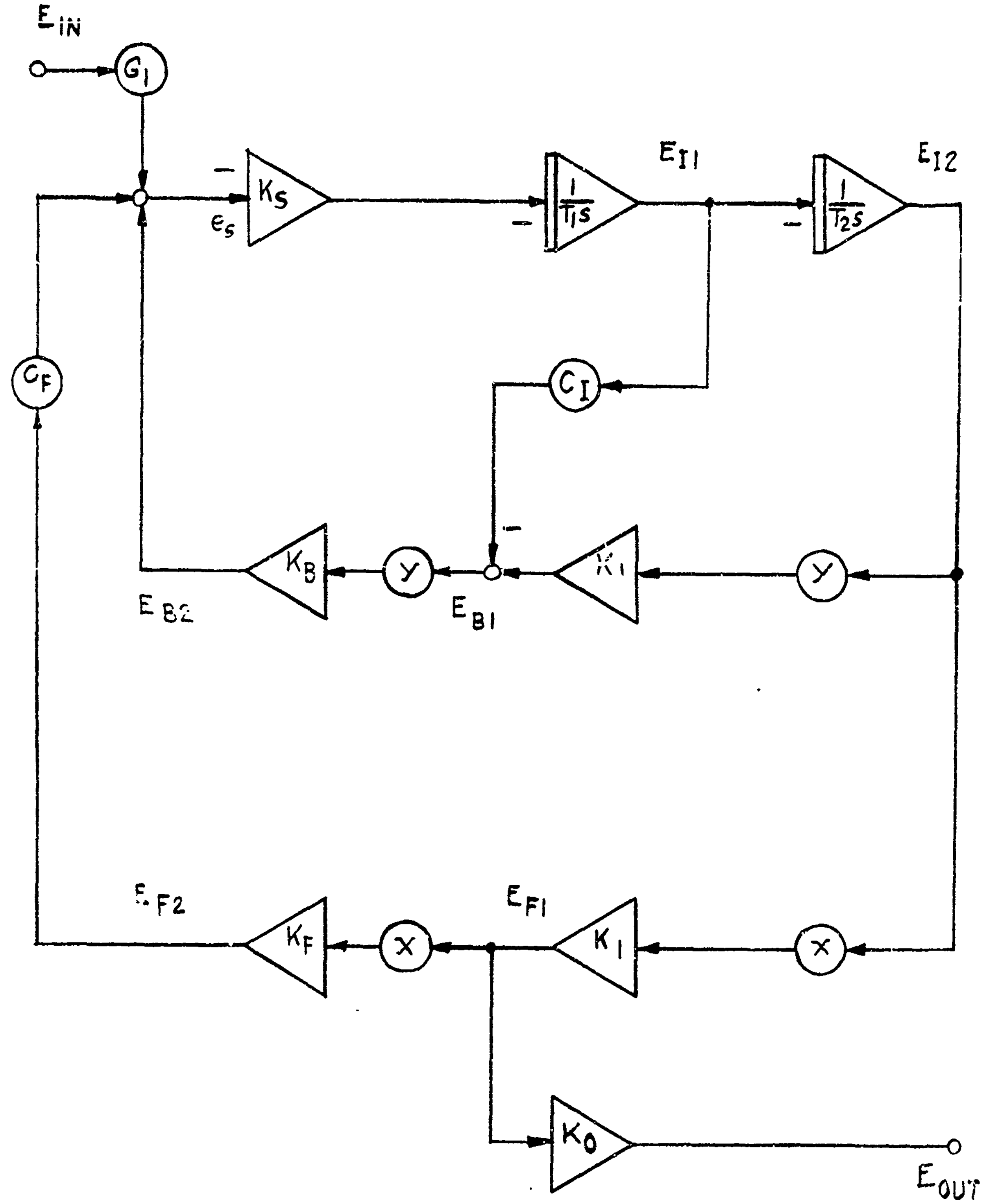


Figure 5. Simplified Transfer-Function Diagram of Resonance Unit

From the last three equalities we obtain the following design constraints:

$$K_s K_B C_I / T_1 = 2\pi F$$

$$K_1 / C_I T_2 = \pi F/2$$

$$K_s C_F K_1 K_F / T_1 T_2 = 4\pi^2 F^2$$

For this design we utilize $F = 10$ KHz, $1/2\pi T_2 = 1$ KHz, $T_1 = 2T_2$, and $C_I = 2$. From these values we obtain:

$$K_s K_B = 10$$

$$K_1 = 5$$

$$K_s C_F K_F = 40$$

The remaining design values selected are $K_s = 4.54$ and $K_s C_F = 2/3$.

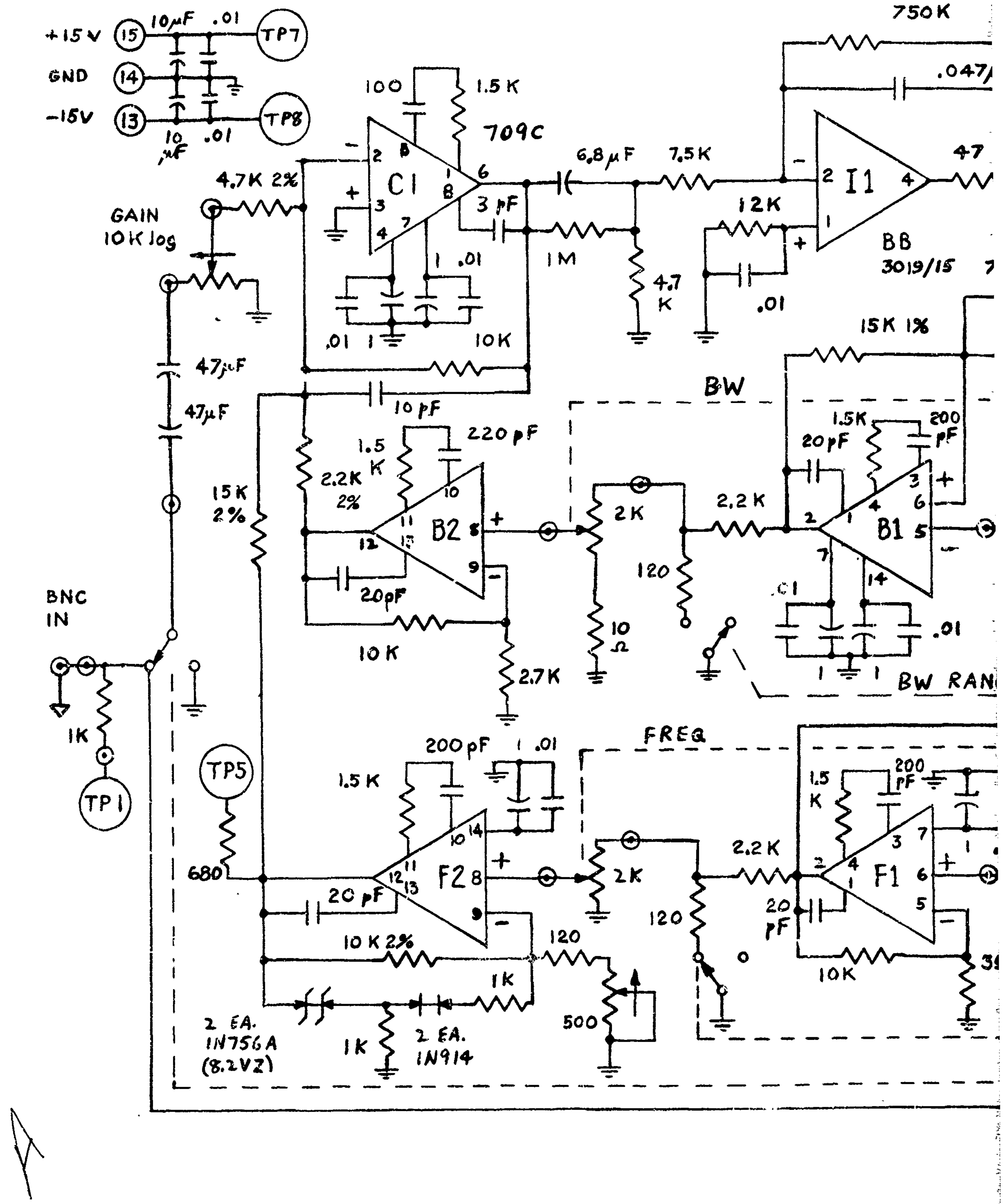
Now that the dynamics of the transfer function are accounted for, by realizing the terms of the denominator, the numerator may be considered and thus the range of gain G_1 required.

If the output were taken at E_{I2} the configuration of Figure 5 would be simpler. One would then expect G_1 to track the main feedback gain $C_F K_F K_I^2$ in order to approach unity closed-loop low-frequency gain. Because x^2 varies over a 2500:1 range, the feedback gain is selected to vary both above and below unity in order that the maximum allowable signal at e_s and E_{I2} should not be unduly restricted. This limits the reduction of signal-to-noise ratio occurring at the extreme f dial settings. Selection of the output configuration shown strikes a balance between good overall dynamic range, close coupling to the common E_{I2} signal, and reduction of the necessary range of the gain control.

The gain control uses an infinite-resolution logarithmic potentiometer for fine manual adjustability. It is passively connected to the input such that the input impedance varies over the restricted range of 3K to 10K. The negative sign in the low-frequency closed-loop transfer function results from the design configuration and implementation.

Figure 6 shows a schematic diagram of the Resonance Filter. Voltage divider networks have been used to prevent loading and to avoid emitter followers. Direct local range switching is used at low impedance levels.

Two circuit additions introduce a minimum damping, and therefore a minimum bandwidth sufficient to insure stability when bandwidth is set near zero (highest Q) but not so large as to affect the desired range of operation.



+15V
GND
-15V
15
14
13
10µF
.01
10 µF
TP7
TP8
100
1.5K
709C
C1
4.7K 2%
GAIN
10K log
3 pF
6.8µF
7.5K
1M
4.7 K
12K
.01
750K
.047
I1
BB
3019/15
47
10K
47µF
47µF
10 pF
220 pF
1.5 K
2.2K 2%
15K 2%
B2
20pF
10 K
2.7K
BW
2K
120
10 Ω
2.2 K
15K 1%
1.5K
200 pF
20 pF
B1
.01
BW RAN
BNC IN
1K
TP1
TP5
680
FREQ
200pF
1.5 K
F2
20 pF
10 K 2%
120
500
1K
2 EA. 1N756A (8.2VZ)
2 EA. 1N914
F1
10K

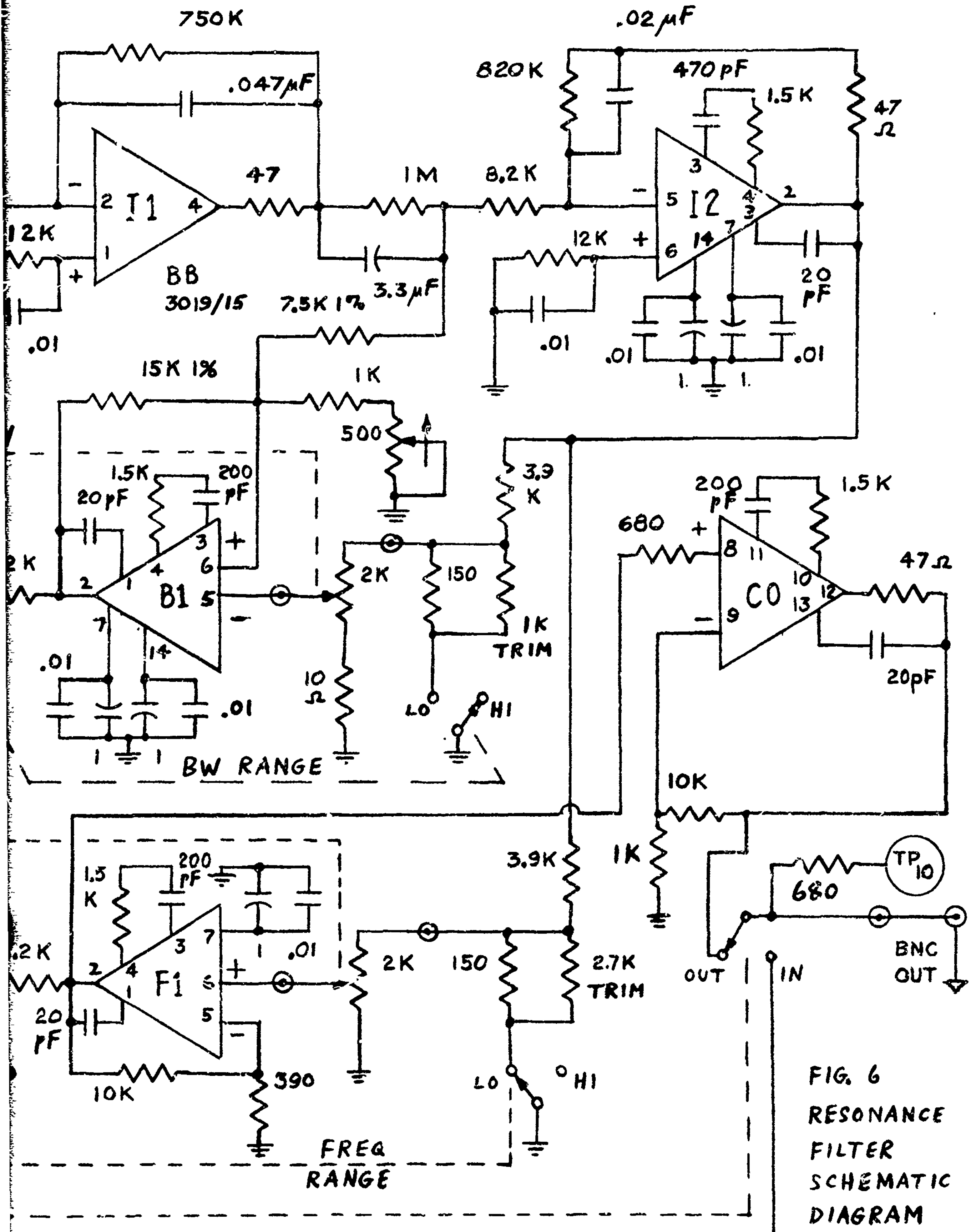


FIG. 6
RESONANCE
FILTER
SCHEMATIC
DIAGRAM

B

First, a bias-stabilizing feedback resistor in each integrating amplifier modifies its transfer function from $1/Ts$ to $1/T(s + 0.01/T)$. Since this is the only frequency-dependent portion, the entire transfer function is modified from H (s) as given above to H $(s + 0.01/T)$. The effect of this translation in the complex frequency domain is multiplication of the corresponding time function h (t) to produce h (t) $E^{-0.01t/T}$. The inverse transform of H_1 (s) yields the impulse response h_1 (t), which has the form:

$$h_1\ (t) = \frac{b^2 + a^2}{a}\ e^{-bt} \sin at$$

Thus the effect of this modification on the dynamic response is to increase $b = \pi yF$ by an amount $0.01/T$, so that y is effectively increased by $0.01/\pi FT$. This has a net value of 0.0015 (10t) = 0.015 turn of dial offset.

The second circuit addition is the insertion of a small resistor to ground at the lower end of each bandwidth potentiometer section. This raises the minimum bandwidth by an additional amount of 10t/200 = 0.05 turn, and further improves the built-in dynamic stability. The total offset of 0.065 turn is calibrated out by setting the dial to read 0.065 turn at pot electrical zero. This corresponds to 6.5 Hz minimum achievable bandwidth on the low range, which is well below the 20 Hz minimum bandwidth setting required.

Adequate bias stabilization of the integrating amplifiers is provided by local feedback when the gain of the main feedback loop is low. At the higher main loop feedback gains, however, tighter bias stabilization is needed which is accomplished using negative DC feedback around the main loop. Because of the cascade arrangement, a Burr-Brown Model 3019/15 low-input-current, low-noise amplifier is used in the first integrator, instead of the 709-type amplifiers used elsewhere.

Two circuit precautions protect the stabilizing signal of the inner (bandwidth) feedback loop from being attenuated or phase-shifted in the summing amplifier due to excessive output from the outer (frequency) feedback loop under overload conditions. The frequency-signal gain factor K_SC_F (See Figure 5) is set below unity, and a soft limiter is provided at the output of Stage F2. These precautions prevent large-signal nonlinear instabilities that might otherwise be initiated under transient overload conditions.

The ideal response function for the Resonance Filter already includes a roll-off above the tuning frequency, so that no modifications are needed for this purpose. All amplifier stages have closed-loop roll-off corner frequencies greater than 100 KHz.

3.4 Resonance Filter Measured Response

The frequency dial calibration of the Resonance Filter was checked at resonance with low bandwidth at various frequency settings on both ranges by comparison with a frequency counter, and agreement well within the ± 2% specification was obtained.

The bandwidth dial calibration was checked quickly with moderate accuracy by setting a fixed signal frequency and tuning both dials to obtain unity gain and 90-degree phase shift on the oscilloscope. This condition corresponds to a second-order network with 0.5 of critical damping tuned to its natural (undamped) resonant frequency. The expressions for this relationship are:

$$0.5 = b/\sqrt{a^2 + b^2}$$

$$2 f_s = \sqrt{a^2 + b^2}$$

When these are solved, the relations given previously yield the results $BW = f_s$ and $f = BW\sqrt{3}/2$. The corresponding damped Q of $\sqrt{3}/2 = 0.866$ for this chosen test condition is slightly beyond the Q = 1 boundary of the required tuning range, providing greater than normal sensitivity to the bandwidth value.

The measured settings shown in Table 3 differ from the ideal values by less than the ± 10% bandwidth calibration specification. Measurement accuracy is limited by the use of the oscilloscope face for the relative gain and the Lissajous phase-shift pattern.

Table 3. Resonance Filter Bandwidth Measurements

Signal Frequency f_s Hz	Dial	Measured Setting turns	Ideal Setting turns	Error %
2K	BW	1.96	2.00	-2
	f	1.71	1.73	-1
1000	BW	9.24	10.0	-8
	f	9.00	8.66	4

3.5 Resonance Filter Maintenance and Calibration

The remarks in Section 2.5 concerning the limited circumstances under which recalibration of the Null Filter might be required also apply to the Resonance Filter.

The following Factory Calibration Procedure utilizes a DC digital voltmeter to set 10:1 attenuation ratios within 0.2% accuracy. AC gains should be set to unity within 2% accuracy, utilizing either an analog or digital AC voltmeter. Refer to the schematic of Figure 6.

Factory Calibration Procedure

1. Check alignment of electrical zero of each BW pot section referred to its lower terminal, which corresponds to 0.065 turn dial reading. Check electrical zero of each FREQ pot section, corresponding to zero dial reading.

2. Check for unity gain magnitude from I1 output to I2 output, when a 1000 Hz input signal is applied to the unit. Trim I2 input resistor or feedback capacitor if necessary.

3. Check for unity gain from I2 output to B1 output with 50 Hz input signal and BW set at 2 KHz. Adjust B1 potentiometer as needed.

4. Check for unity gain from B1 output to B2 output with BW set at 4.54 KHz. Trim B2 feedback resistor if necessary.

5. Check that the transmission from I2 output to the high end of the B1 section BW pot falls by 10:1 with low BW range switching. Remove I2 and apply DC voltage with BW and f at minimum dial settings. Trim 150 ohm resistor if necessary.

6. Check that the transmission from B1 output to the high end of the B2 section BW pot falls by 10:1 with low BW range switching. Remove B1 and apply DC voltage. Trim 120 ohm resistor if necessary.

7. Apply 4000 Hz input signal, set BW to minimum and f to 4 KHz (high range), and vary f dial to find peak response of unit. Adjust F2 potentiometer to obtain f = 4 KHz dial setting for peak response.

8. Apply 900 Hz input signal with BW set to minimum, set f to 900 Hz (low range), and vary f dial to find peak response. Trim 150 ohm resistor at F1 if necessary to obtain f = 900 Hz dial setting for peak response.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

SIGNATRON, Inc.
594 Marrett Road
Lexington, Massachusetts 02173

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

INVESTIGATION OF FACTORS AFFECTING THE QUALITY OF VOCODER SPEECH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final - April 15, 1967 through May 17, 1969 (Approved 17 June 1969)

5. AUTHOR(S) (First name, middle initial, last name)

Thomas H. Crystal

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May /7, 1969	78	13

8a. CONTRACT OR GRANT NO.
F19628-67-C-0292

b. PROJECT NO. Task, Work Unit Nos.
4610-02-01

c. DoD Element: 62405454

d. DoD Subelement: 674610

9a. ORIGINATOR'S REPORT NUMBER(S)

SIG-CR-594

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

AFCRL-69-0256

10. DISTRIBUTION STATEMENT

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

11. SUPPLEMENTARY NOTES

TECH, OTHER

12. SPONSORING MILITARY ACTIVITY

Air Force Cambridge Research Laboratories (CRB)
L. G. Hanscom Field
Bedford, Massachusetts 01730

13. ABSTRACT

Research into and the development of instrumentation for the investigation of factors affecting the quality of vocoded speech are documented. The work reported was specifically concerned with developing a better understanding of the role of the vocal source in the production of both synthetic speech and of natural speech. The design of and operating instructions for the VOTIF vocal track inverse filter - built as part of the program - are presented. A theoretical determination of the interaction between the vocal source and vocoder channel filters has been made and the effect of spectrum flattening on the peak factor and power of a vocoder channel have been computed. Lastly, the pulsed excitation of resonances is discussed. A form of pitch jitter which could either maximize vocal output or minimize vocal tract impedance effects is reported on.

DD FORM 1473 1 NOV 65

Unclassified
Security Classification

SIGNATRON, Inc.
research and consulting
area code 617 • TEL. 862-3365
MILLER BUILDING • 594 MARRETT ROAD • LEXINGTON, MASSACHUSETTS 02173

AD 690179

ERRATA SHEET

AFCRL-69-0256

INVESTIGATION OF FACTORS AFFECTING
THE QUALITY OF VOCODER SPEECH

May 17, 1969

page 3-8 In the first line of Eq. (3.13a),

$; \rho = \hat{n}$ should be $; \rho \neq \hat{n}$

page 3-8 In the first line of Eq. (3.13b),

$; \rho = \hat{n}$ should be $; \rho \neq \hat{n}$

Next to last page (Form 1473), item 6, Report Date

May 7, 1969 should be May 17, 1969.

Last page (reverse of Form 1473), item 14, Key Words should be added as follows:

Inverse Filter
Instrumentation Speech Analysis
Speech
Speech Production
Vocal Source
Vocal Tract Inverse Filter
Vocoder
Vocoder Excitation
Vocoder Filter
Vocoder Quality

Ref: AD 690179